UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Iron Man(ipulation): unravelling the iron responsive genes' network in malaria parasites

Miguel Dinis Monteiro dos Santos

Mestrado em Bioinformática e Biologia Computacional
Especialização em Biologia Computacional

Dissertação orientada por:
Dr. Ksenija Slavic
Dr. Octávio Paulo

2016

Acknowledgments

It would not be possible to finish this thesis without the help and support from several people. I will only mention the ones that were closer to me during this process, but I am truly thankful to all of you that contributed directly and indirectly to the outcome of this work.

I am truly thankful to Dr. Maria Mota for accepting me in her lab and allowing me to perform my Master thesis in her group. I want to thank to my internal supervisor, Dr. Octávio Paulo for all the support and confidence. To Ksenija, for being the best supervisor I could ever ask. Since the first day I felt your trust in my work and that you were always listening to my ideas and suggestions. At the same time, you were always frontal with me, teaching me in the best way you can, by correcting my mistakes and congratulating me when things were properly done. You will be truly important for everything I will do in my life after this, and I won't never forget your advices (both personal and scientific).

To Torcato Martins and Filipa Domingos, for providing me with one of the best summer I had and being a major part of my scientific and personal growth.

To everyone in MMota lab, for the constant help and support, as well as the friendship and funny moments. Here I have to highlight some names: thank you Ângelo, Vanessa, Iset, Inês B. and João V., for sharing their knowledge, advices and for the challenging conversations, Liliana for all the help and enthusiastic teachings, to Jon for making every night The Night and to Maria Inês Marreiros for the unconditional friendship and support. If our lab is a family, for sure you are my sister.

To everyone in LFigueiredo and MPrudencio labs, for the help and friendship. I want to give a special thanks to Daniel Neves for the advices, the productive brainstorming and constant availability for help and to Fábio for the friendship. I am also thankful to António Mendes for the interesting discussions. Thank you to Ana Rita Grosso for the advices and help provided. I want to give a special thanks to the STAB VIDA staff for the good work and support.

To all my friends, for being the best I have and for being always there with all the patience and friendship: to Nina, one of my oldest friends, to Pedro for the nights and conversations and to Camila for the infinite patience and constant care; you are one of the people I wish more the best. To Chico and Carolina, two of my most recent friends without whom this year wouldn't be the same: Chico for the deep, fun and honest friendship, you became like an older brother to me; Carolina for becoming one of my best friends and one of the most important people during this process, watching you overtake this year was truly inspirational. To my great friend João Simão: despite the distance I felt that you were always close. You are one of the people I respect and I am more proud of. To Maria Beatriz, my best and oldest friend. Every day you keep surprising me and make me feel lucky for met you. To Malhadas, my best friend: I know you would give an arm and a leg for me and I would do the same for you. I miss you and I am proud of you every day, I know things with you will never change.

To Marco, one of the most brilliant people I ever met and one of my best friends. For all the conversations, nights and honest advices (if some were difficult to hear, I cannot imagine how difficult they were to tell). I trust you with all my heart and you are one of the few people I feel I can be completely myself. Your unconditional friendship and support during the last years were indescribable and completely fundamental. Thank you for challenging me and make me break my limits. Thank you for being always frontal with me. You are an amazing person and I know you will be incredible in everything you do.

To Filipe, Bia, Inês (both) and Patrick for more than living with me, for being my home: you were part of everything during the last 5 years and I could not have been luckier for meeting you. You make me leave and enter at home with a smile every day and calling you family is not enough. We already came through so many, and there is nothing that I didn't tell you already. I love you with all my heart. To everyone that passed through our house: James, Afonso, Dani and Alice. To James (the cat) and Luke for being partners in crime and friends for life.

To my Family, to whom I owe the person I am and everything I got and achieved. I could not be happier with you. To my lovely Mother, the woman of my life and the one that was always there. I do not have words to explain how much I love you and feel blessed for being your son. I wouldn't change anything about you. To my lovely Father, one of the two best man I know. I hope one day I will be half the man you are. Thank you for teaching me what is truly important and for making me feel the best. To Tiago, my big brother: you are the other best man I know, my role model and a friend for life. I would not have achieved anything I did without your help, advices and unconditional and tireless support, especially during this year. Part of this work I owe to you, as well as most of what I am. Thank you for having your home always available for me, and for loving me as you do. You are the best brother I could ever ask. To Maria, my little sister and the joy of my eyes. You are such an inspiration! I hope I can be in every step of your life and you know that you can count with me for all your life. I love you since the day you were born. To my little babies Matilde and Simão for making me feel so loved, and to Maria Helena for loving me as a son. To my grandparents which would give everything for me, and all my uncles, aunts and cousins. I am the luckiest person ever.

To the most important person ever, to whom I already told everything I could and that were part of every step. Diana, every day I try to find a way to show you how much I love you and how much I am thankful to you. You are the person I respect and are more proud of and I keep learning with you every day. I have never met someone like you. I know with all my heart that you are the only person I will ever love. You never failed to me, and were by my side in every moment. I hope you will continue being part of my life as you are until the last day. I will always love you and I know we don't need to tell anything to each other in order to know what we are feeling. Thank you my love, for being as you are. Thank you!

Resumo

A Malária é uma doença infeciosa causada por várias espécies de parasitas protozoários do género *Plasmodium.* Entre eles, *P.* falciparum é o que provoca a maior mortalidade em humanos. O parasita apresenta um ciclo de vida complexo, alternando entre o inseto vetor que transmite a forma infeciosa do parasita e o hospedeiro mamífero, no qual após a infeção o parasita passa por uma enorme replicação causando os sintomas da doença. Após a picada de um mosquito infetado, formas móveis denominadas esporozoítos são injetadas dentro do hospedeiro mamífero e viajam até ao fígado onde invadem hepatócitos. Após replicar dentro dos hepatócitos, milhares de novos parasitas são libertados para a corrente sanguínea onde invadem os glóbulos vermelhos. Durante a fase sanguínea da infeção o parasita passa por diversas formas de desenvolvimento, utilizando um programa de expressão génica altamente coordenado onde genes codificantes de proteínas para funções específicas de cada forma são ativados apenas quando são necessários. Ocasionalmente, algumas formas da fase sanguínea diferenciam-se em gametócitos, a forma sexual transmissível para os mosquitos. Durante este complexo ciclo de vida, para assegurar a sua sobrevivência e replicação, o parasita necessita de sequestrar nutrientes e micronutrientes de várias fontes dos tecidos do mosquito e do mamífero. Portanto, foi posta a hipótese de que os parasitas causadores de malária possuem mecanismos adequados de adaptação de maneira a possibilitar a utilização desta enorme variedade de recursos. O ferro é um dos micronutrientes mais importantes, estando envolvido em diversos processos celulares essenciais para a sobrevivência, como a replicação do ADN e a produção de ATP. Por outro lado, o ferro é altamente tóxico quando se encontra em excesso, o que implica que a sua concentração dentro das células tenha que ser mantida num intervalo restrito. Isto é tipicamente alcançado através da regulação da expressão de proteínas envolvidas na aquisição de ferro, armazenamento e metabolismo. Apesar do papel fundamental do ferro e da sua regulação para a sobrevivência celular, uma grande quantidade de informação básica encontra-se em falta no que diz respeito à forma como o parasita lida com este valioso recurso. De maneira a abordar esta questão, realizámos uma análise do transcriptoma total de *P. falciparum* durante a fase sanguínea em condições a curto prazo de privação de ferro e exposição a excesso de ferro. Ao identificar os genes diferencialmente expressos entre as diferentes condições e os respetivos controlos, pretendemos encontrar genes envolvidos na deteção e regulação de ferro. Os genes identificados como diferencialmente expressos possuem uma grande variedade de funções, desde fatores de transcrição e proteínas de ligação a ARN mensageiro até componentes do sistema ubiquitina-proteossoma, ARNs não codificantes, transferases de grupos metilo de histonas, vários transportadores e proteínas envolvidas no metabolismo. Estes resultados sugerem a existência de diversos níveis de regulação dos genes relacionados com o ferro, tanto ao nível da transcrição e da pós-transcrição, bem como ao nível da degradação de proteínas e mecanismos epigenéticos, indicando uma rede altamente complexa envolvida na manutenção da homeostasia do ferro em *Plasmodium*.

Palavras-chave – Ferro, *Plasmodium falciparum*, Transcriptoma, RNA-seq

Abstract

Malaria is an infectious disease caused by several species of protozoan parasites of the *Plasmodium* genus. Among them, *P. falciparum* is the one causing the greatest mortality in humans. The parasite has a complex life-cycle, switching between the insect vector which transmits the infective form of the parasite and the mammalian host, in which upon infection parasite goes through enormous replication causing the symptoms of the disease. Upon the bite of an infected mosquito, motile forms called sporozoites are injected into the mammalian host which travel to the liver and invade hepatocytes. After replicating inside hepatocytes, thousands of new parasites are released into the bloodstream where they invade red blood cells. During the blood stage of the infection the parasite goes through several developmental stages, using a tightly coordinated gene expression program where genes encoding proteins for stage-specific functions are activated only when they are required. Occasionally, some blood stage forms differentiate into gametocytes, the sexual stage transmissible to the mosquitoes. During this complex life cycle, to support its survival and replication, parasite needs to scavenge nutrients and micronutrients from various sources of the mammalian and insect tissues. Thus, it is hypothesized that malaria parasites must have adequate mechanisms of adaptation to be able to use this wide range of resources. Iron is one of the most important micronutrients, as it is involved in several cellular processes essential for survival, such as DNA replication and ATP production. On the other hand, iron is highly toxic when found in excess, therefore its intracellular concentrations need to be kept in a very narrow range. This is typically achieved through regulation of expression of proteins involved in iron acquisition, storage and metabolic usage. Despite the critical role of iron and its regulation for cell survival, a great deal of basic information is missing regarding the way that malaria parasites deal with this valuable resource. In order to address this question, we performed genome-wide transcriptome analysis of *P. falciparum* blood-stage by RNA-seq in conditions of short-term iron deprivation and exposure to iron excess. By identifying the differentially expressed genes between the different conditions and the respective control, we aimed to find genes involved in iron sensing and regulation. The genes identified as differentially expressed had a great variety of functions, from transcription factors and mRNA-binding proteins to components of the ubiquitin-proteasomal system, ncRNAs, histone methyl-transferases, several transporters and metabolism-related proteins. These results suggest the existence of several layers of regulation of the iron-responsive genes, both at the levels of transcription and post-transcription, as well as protein degradation and epigenetic mechanisms, pointing to a highly complex network involved in the maintenance of iron homeostasis in *Plasmodium*.

Key-words – Iron, *Plasmodium falciparum*, Transcriptome, RNA-seq

Resumo alargado

A Malária é uma doença infeciosa causada por várias espécies de parasitas protozoários do género *Plasmodium*. Apesar dos esforços para a prevenção, da redução no número de casos e do investimento na investigação de novas terapias e vacinas anti-maláricas, esta doença continua a representar um grande fardo económico nas áreas endémicas. Entre as espécies existentes, *P. falciparum* é aquela que causa maior número de casos e maior mortalidade em humanos. Estes parasitas apresentam um complexo ciclo de vida, que se inicia quando uma fêmea de mosquito do género *Anopheles* (o hospedeiro invertebrado) injecta esporozoítos, uma forma infeciosa altamente móvel, na derme do hospedeiro mamífero (humanos no caso de *P. falciparum*) durante a sua refeição de sangue. Os esporozoítos viajam para a corrente sanguínea e dirigem-se ao fígado, onde invadem hepatócitos dentro dos quais se multiplicam em milhares de merozoítos. Estes viajam para a corrente sanguínea e invadem os glóbulos vermelhos, iniciando-se a fase sanguínea (assexuada) onde os parasitas apresentam um elevado grau de regulação da expressão génica, onde genes com funções específicas para cada forma são ativos apenas no exato momento em que são necessários. Durante a fase sanguínea o parasita assume diversas formas morfologicamente distintas: uma forma denominada anel devido à sua morfologia característica, os trofozoítos, altamente ativos metabolicamente, e os esquizontes. Dentro dos esquizontes formam-se novos merozoítos que irão invadir novos glóbulos vermelhos. Ocasionalmente, alguns merozoítos diferenciam-se em gametócitos feminino ou masculino, a forma sexuada infeciosa para os mosquitos. Quando os mosquitos picam um hospedeiro mamífero infetado, os gametócitos são transportados para o sistema digestivo destes onde vão dar origem aos respetivos gâmetas. O zigoto diploide móvel resultante da fecundação denomina-se oocineto, que invade as paredes do sistema digestivo do mosquito e diferencia-se num oocisto. Dentro deste ocorrem várias divisões mitóticas que darão origem a novos esporozoítos, que viajarão para as glândulas salivares dos mosquitos e serão transmitidos para um novo hospedeiro mamífero durante a próxima refeição de sangue.

Durante este complexo ciclo de vida, os parasitas lidam com diversos recursos do hospedeiro com uma larga gama de concentrações, o que exige que estes possuam um sofisticado sistema de deteção e regulação dos mesmos. Entre estes micronutrientes, o ferro possui uma elevada importância devido ao seu papel central em diversos processos celulares, desde a replicação do ADN e produção de ATP até à participação em diversas reações enzimáticas como cofactor. Este micronutriente possui características de oxidação-redução únicas que lhe conferem o seu papel central em diversas reações bioquímicas. O facto de alternar entre dois estados de oxidação (2^+ e 3^+) permite-lhe associar-se com proteínas, ligar-se ao oxigénio e transportar eletrões, entre outras funções. Devido ao seu papel fulcral, este é altamente regulado na grande maioria dos organismos, desde as leveduras aos mamíferos onde diversos mecanismos de regulação ao nível da transcrição e da pós-transcrição atuam para manter a sua homeostasia, desde fatores de transcrição a proteínas que se ligam a ARN mensageiro. Apesar da sua elevada importância na relação parasita-hospedeiro, os mecanismos de regulação deste recurso em *Plasmodium* são ainda altamente desconhecidos. Exceções são a identificação de dois transportadores de ferro, o facto de o parasita regular a expressão da enzima hepática hepcidina, de extrema importância na regulação do ferro em mamíferos e do envolvimento do heme (grupo prostético da hemoglobina) no modo de ação da artemisinina, um dos medicamentos mais usados para combater a doença. Tendo estas evidências em conta, postulámos que estes mesmos mecanismos de regulação da expressão de genes envolvidos no metabolismo do ferro se encontram conservados no parasita, mantendo os seus níveis altos o suficiente, possibilitando a sua utilização sem que estes se tornem tóxicos.

Para a abordagem destas questões, as técnicas de sequenciação de larga escala de nova geração tornaram-se estratégias cada vez mais comuns e foram já aplicadas a diversos organismos incluindo parasitas do género *Plasmodium*. O facto de o seu genoma já se encontrar completamente sequenciado representa também uma vantagem neste tipo de estudos. Uma das técnicas mais utilizadas é a sequenciação do transcriptoma completo, ou seja, a determinação da expressão génica global de um dado organismo num dado momento denominada RNA-seq. Esta consiste na extração do ARN total de uma amostra biológica, construção de uma biblioteca de sequenciação através da conversão para ADN complementar, ligação de adaptadores que permitem a imobilização num suporte sólido e sequenciação utilizando nucleótidos artificiais que possuem um fluoróforo específico ligado. O sinal emitido será lido pelo sequenciador, permitindo a identificação das bases em cada posição. Esta técnica possibilita a abordagem de diversas questões, sendo a mais utilizada a identificação de genes diferencialmente expressos entre condições (organismo selvagem *vs.* organismo mutante ou situação controlo *vs.* tratamentos). Utilizando esta técnica, caracterizou-se o transcriptoma ao longo da fase sanguínea do parasita, revelando resultados muito interessantes. Ao longo desta fase do desenvolvimento ocorre uma onda de ativação e inibição dos genes específicos para cada forma morfológica e função específica, com cada gene tendo um único pico de expressão durante esta fase apenas quando é estritamente necessário.

Pelos motivos descritos acima, propusemo-nos a identificar novos genes envolvidos na rede de regulação do ferro no parasita e, para tal, levámos a cabo uma experiência de sequenciação de ARN (RNA-seq) em diferentes condições de ferro de forma a identificar os genes diferencialmente expressos, que estarão potencialmente envolvidos neste processo. Para tal desenvolvemos um sistema artificial que introduz alterações na concentração de ferro do ambiente do parasita, nomeadamente uma condição de excesso e de deficiência de ferro. Após confirmarmos que o sistema era funcional e inofensivo para o desenvolvimento e viabilidade dos parasitas, identificámos diversos homólogos de genes envolvidos no metabolismo do ferro em organismos-modelo e testámos se a sua expressão era alterada pelos nossos tratamentos através de PCR quantitativo. Estes resultados validaram o nosso sistema, e permitiram-nos avançar para a experiência de RNA-seq, de maneira a caracterizar o transcriptoma de parasitas submetidos às diferentes condições de ferro. Para tal, utilizámos um protocolo no qual os fragmentos obtidos são alinhados contra o genoma de referência utilizando o software TopHat. De seguida procedemos à assemblagem dos transcritos utilizando o software Cufflinks e, por fim, realizámos a análise de expressão diferencial utilizando o software Cuffdiff. Devido ao facto de os parasitas dos nossos replicados biológicos, realizados em dias distintos, não se encontrarem exatamente na mesma forma morfológica, deparámo-nos com alguns problemas aquando do agrupamento dos nossos resultados. Ao compararmos os nossos dados de transcriptoma com os dados publicados ao longo da fase sanguínea, chegámos à conclusão que aquando da extração de ARN, os parasitas encontravam-se numa forma um pouco mais avançada no segundo dia quando comparando com o primeiro. A melhor solução foi a comparação das diferentes condições dentro de cada dia de experiência, tentando identificar genes com comportamentos semelhantes entre os tratamentos em cada dia.

Os genes identificados como diferencialmente expressos entre as condições de ferro e os respetivos controlos possuem uma grande variedade de funções, desde fatores de transcrição, proteínas que se ligam a ARN mensageiro, transferases de grupos metilo de histonas, enzimas envolvidas na degradação de proteínas pelo complexo ubiquitina-proteossoma, diversos transportadores e até ARNs não codificantes. Estes resultados demonstram que o parasita não mantém a homeostasia do ferro usando apenas uma estratégia, mas através de uma rede de genes e mecanismos altamente complexa e regulada a diversos níveis, o que permite a continuação deste projeto em diversas direções altamente interessantes.

Table of Contents

List of tables and figures

List of abbreviations and symbols

RBCs	Red Blood Cells
IDC	Intraerythrocytic Developmental Cycle
DNA	DeoxyRibonucleic Acid
Fe-S	Iron-Sulfur
ROS	Reactive Oxygen Species
GSH	Glutathione
ER	Endoplasmic Reticulum
mRNA	messenger Ribonucleic Acid
ARE	AU-rich Element
UTR	Untranslated Region
TCA	tricarboxylic acid
DMT1	Divalent Metal Transporter 1
TfR1/2	Transferrin Receptor 1/2
IRP1/2	Iron Regulatory Protein 1/2
IRE	Iron Responsive Element
ZIPCO	ZIP domain-containing protein
VIT	Vacuolar Iron Transporter
CRT	Cyclic Reversible Termination
µ	micro
M	Molar
BLAST	Basic Local Alignment Search Tool
MCM	Malaria Complete Medium
DFO	Deferoxamine
qPCR	quantitative PCR
RIN	RNA integrity number
GTF	Gene Transfer Format
FPKM	Fragments Per Kilobase of transcript per Million mapped fragments
MDS	MultiDimensional Scaling
FDR	False Discovery Rate
SERA	Serine Repeat Antigen
lncRNA	long non-coding RNA
GO	Gene Ontology
ATP	Adenosine Triphosphate
HIF	Hypoxia Inducible Factor
ncRNA	non-coding RNA

1 - Introduction

1.1 – The disease and the parasite

Malaria is an infectious disease responsible for approximately 500,000 deaths and 200 millions of infections worldwide only in 2015 (Anon 2015). Despite the progress in research and the decrease in the number of cases worldwide, malaria is still a major health and economic concern in endemic areas. The disease is caused by several species of the protozoan parasite *Plasmodium*, which are transmitted to the hosts through the bite of infected *Anopheles* mosquitoes. Five *Plasmodium* species are infectious to humans: *P. falciparum, P. vivax, P. ovale, P. malariae and P. knowlesi*. Among these, *P. falciparum* is causing the greatest mortality (Anon 2015). Since for obvious reasons, *in vivo* studies in humans are complicated, species infectious to rodents like *P. berghei, P. yoelii and P. chabaudii* are widely used in malaria research and of major importance to the understanding of the parasite's biology as well as disease's pathology (Craig et al. 2012).

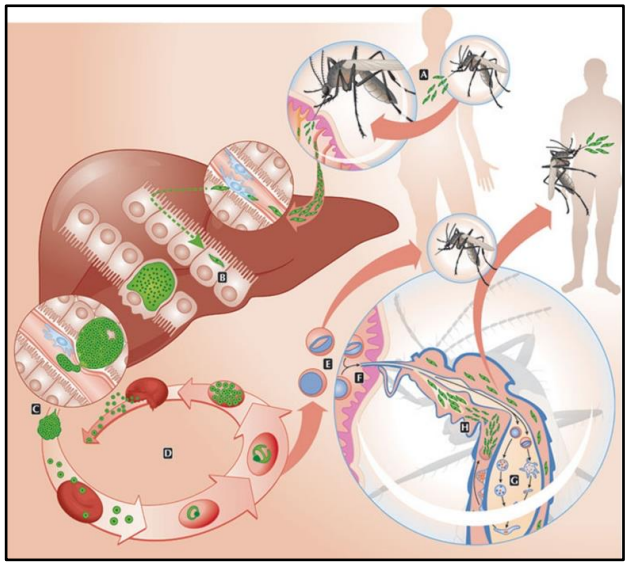


Figure 1.1 - *Plasmodium* life cycle: A - female *Anopheles* mosquitoes infected with sporozoites injects them in the mammalian host dermis. **B -** After reaching the bloodstream, the sporozoites travel to the liver where they invade hepatocytes and **C -** replicate into thousands of merozoites, which will be released into the bloodstream and infect red blood cells (RBCs). During the blood stage, the parasite goes through three different stages: ring, trophozoite and schizont. **D -** During the schizont stage, new merozoites are formed which will reinvade RBCs. **E -** The sexual stage begins when some merozoites develop into male or female gametocytes, the form infective to mosquitoes. **F -** Inside the mosquito the gametocytes differentiate in the respective gametes and inside the mosquito gut a zygote is formed, which by itself differentiates into a ookinete. **G -** The ookinete invades the midgut wall and develops into an oocyst, where several mitotic divisions take place, producing new sporozoites. **H -** This sporozoites migrate to the mosquito salivary glands, and are ready to infect a new mammalian host during the mosquito's blood meal. Figure from (Portugal et al. 2011)

The life cycle of *Plasmodium* begins when during a blood meal, a female *Anopheles* mosquito infected with the parasite injects sporozoites (motile, spindle-shaped asexual cells) in the host dermis (Fig. 1.1-A). Inside the mammalian host, the parasite goes through a complex life cycle. After reaching the bloodstream, the sporozoites travel to the liver where they invade hepatocytes (Fig. 1.1- B). Inside the hepatocytes each parasite will develop and replicate into thousands of new merozoites, that will be released into the bloodstream and infect red blood cells (RBCs) (Fig. 1.1-C). This step initiates the intraerythrocytic developmental cycle (IDC), in which the parasite goes through three different stages: ring (named after its characteristic morphology), trophozoite and schizont. In the trophozoite stage, the host cytoplasm is ingested and haemoglobin is proteolysed into amino acids. In the schizont stage, the parasite prepares for reinvasion of new RBCs by replicating DNA and forming up to 16 to 32 new merozoites (depending on the species) through the process of schizogony (Fig. 1.1-D). The blood stage is responsible for the disease's onset and symptoms (Miller et al. 2013) and is also the target for the vast majority of antimalarial drugs. The IDC duration differs between *Plasmodium* species, being 24 hours for *P. berghei* and 48 hours for *P. falciparum*, for example. *Plasmodium* enters a sexual stage when some merozoites in the erythrocytes develop into male or female gametocytes (Fig. 1.1-E), which upon

transmission to mosquito differentiate into male and female gametes, respectively (Fig. 1.1-F). Inside the mosquito's gut the resulting diploid zygote will differentiate into an ookinete (a motile zygote). The ookinete invades through the midgut wall of the mosquito and develops into an oocyst (Fig. 1.1-G). Within the oocysts, several mitotic divisions take place, producing new sporozoites. These sporozoites migrate to the mosquito's salivary glands and from there are injected in the bloodstream of the mammalian host during the next blood meal, starting the cycle again (Figure 1-H.1) (Portugal et al. 2011).

Across this complex life cycle the parasite must be adapted to a variability of nutrient resources, which represents a major challenge. Specifically concerning the micronutrients, such as iron, a tight balance must be achieved to meet the cellular demands but avoid toxicity. Thus, the parasite needs to maximize the usage of the host's micronutrients without compromising its survival. To achieve this narrow homeostasis, it is expected that *Plasmodium* parasites, like other organisms, have mechanisms to sense, acquire, detoxify and metabolize iron. The complex and continuous host-parasite battle for iron has been an attractive research area, given the importance for the outcome of the disease.

1.2 – Pathogen-host battle for iron

It is not possible to talk about life without talking about iron. Its involvement in essential biochemical processes inside cells makes it one of the most important elements for almost all living organisms (Wang & Pantopoulos 2011; Aisen et al. 2001). In cells, iron can be found in the reduced ferrous (Fe^{2+}) and oxidized ferric (Fe^{3+}) forms. When associated with proteins, this flexible coordination and consequent redox reactivity allows it to bind oxygen, transport electrons and be part of enzymatic reactions as a cofactor on its own or as part of prosthetic groups like haem or iron-sulfur (Fe-S) clusters (Wang & Pantopoulos 2011). As such, iron is essential for the process of DNA replication and cellular respiration, among others.

As essential for both the malaria parasite and its mammalian host, iron is one of the factors that contributes for the complexity of the host-pathogen relationship. Apart from two microorganisms known to exceptionally use manganese instead of iron (non-pathogenic lactobacilli and pathogenic *Borrelia burgdorferi*), all bacteria, fungi and protozoa require iron for survival and replication (Weinberg 2009). Thus throughout evolution, both host and pathogen have been contestants in an arms race for the control of this valuable resource (Weinberg 2009): (1) In one hand, hosts developed strategies to keep their iron reservoirs unavailable to their invading opponents, (2) as on the other hand pathogens developed tools to steal it from their hosts. Notably, this host's property is still one of the most evolutionarily conserved innate mechanisms against infection (Drakesmith & Prentice 2012).

The involvement of iron in pathogens' growth was first described in 1944, when Schade and Caroline observed that raw egg white had an inhibitory effect in the growth of certain bacteria. This effect was only reversed when the medium was supplemented with iron, and iron alone was enough to overcome it (Schade & Caroline 1944). Now it is known that this inhibitory effect is due to ovotransferrin, an iron-binding protein present in egg white. In fact, one of the main mechanisms employed by hosts to block iron access to pathogens is by binding it to proteins, like transferrin and ferritin. One of the most common pathogens' counter attack is the use of receptors that bind these proteins (and even haemoglobin directly). Usually this kind of mechanism is host specific (Weinberg 2009). A mechanism that can affect a broader range of hosts is the use of siderophores (from the Greek "iron carrier"). These low molecular mass iron-binding agents scavenge iron and transport it inside cells of bacteria and fungi (Neilands 1995). Again, some hosts developed innate immune responses against siderophores like the lipocalin 2

present in mammalians, that binds to enterocholine, a siderophore of enterobacteria (Flo et al. 2004). Additional mechanism is provided by the mammalian liver hormone hepcidin which acts as master regulator of dietary iron absorption and iron release from macrophages resultant from the recycling of senescent erythrocytes (mechanistic details in *Iron sensing and regulation in mammalian cells*). Also, hepcidin production is induced during infections, reducing the iron available for pathogens (Hentze et al. 2010).

One of the most interesting findings regarding *Plasmodium* and iron, is the ability of blood-stage parasites to stimulate the production of hepcidin in a density-dependent manner, thus inhibiting subsequent liver stage infections (Portugal et al. 2011). As described above, the increase in hepcidin's levels leads to the degradation of ferroportin, causing a redistribution of iron away from hepatocytes where it is needed for the development of liver stage forms. Altogether, these findings reveal two interesting characteristics of the parasite: 1) the ability to alter the expression of a host's hormone, 2) and the presence of cross-stage competition, using iron depletion as a way to inhibit growth (blood *vs.* liver stage parasites).

The examples described above, as well as many more that could be referenced, clearly place iron as a central mediator of host-pathogen interactions, as both parts of the equation continuously evolved and adapted in order to win the battle for iron. However, even though essential, iron can be highly toxic, as in aerobic conditions it promotes the formation of reactive oxygen species (ROS) and generation of highly reactive radicals through Fenton reaction (Aisen et al. 2001). Thus, organisms have evolved sophisticated strategies to sense, acquire, metabolize, and detoxify iron, fulfilling their metabolic needs while keeping its levels not harmful. These processes are extensively studied from the cellular level in organisms as simple as yeast to the systemic levels in mammals, but remain poorly investigated in case of *Plasmodium* parasites. In order to start investigating the possible mechanisms that malaria parasites may use to sense the iron levels and achieve homeostasis, it is necessary to understand how is this achieved in other organisms and investigate any possible evolutionary conservation of these pathways in *Plasmodium.*

1.3 – Yeast iron sensing and regulation

Being the simple single-cell eukaryote, amenable to genetic modifications, has made the budding yeast *Saccharomyces cerevisiae* one of the most attractive models to study the cellular and molecular basis of iron regulation. The regulation of iron metabolism in yeast is achieved through the first line of regulation at the transcriptional level and the second layer of post-transcriptional regulation. Key players of the iron regulation are three transcription factors: two sensing low iron levels (Aft1 and its paralogue Aft2), and one sensing high iron levels (Yap5). Aft1/2 regulate the expression of proteins involved in iron transport at the plasma membrane, vacuolar iron transport and iron metabolism, which combined effect will lead to the transport of iron into the cytosol from both intra and extracellular sources (Outten & Albetel 2013). On the other hand, Yap5 regulates the expression of proteins involved in iron detoxification, namely the vacuolar iron importer CCC1 and the iron-sulfur cluster binding proteins Tyw1 and Grx4 (Li et al. 2011).

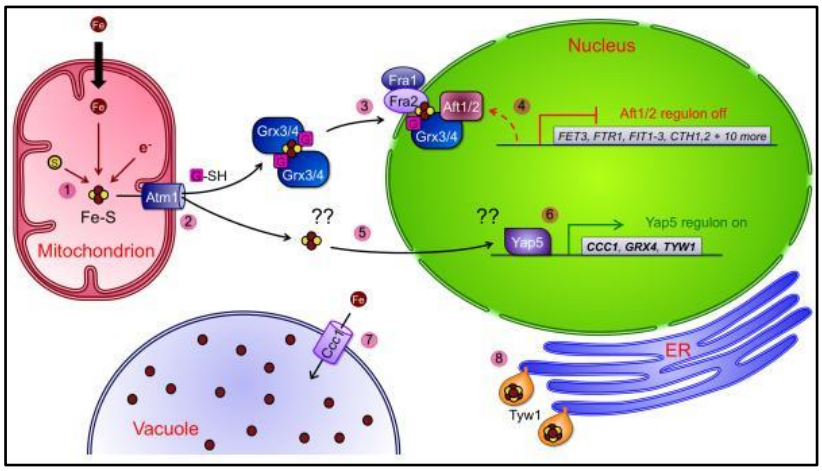


Figure 1.2 - Transcriptional iron regulation in yeast: when iron levels are high, iron-sulfur clusters formed inside mitochondria travel to the nucleus bound to monothiol glutaredoxins Grx3 and 4. Inside the nucleus they bind to the transcription factors Aft1/2, leading to their dissociation from the respective target genes. Additionally, Yap5 is activated and consequently expression of its target genes, namely the iron detoxifier CCC1, GRX4 and ER-bound protein TYW1. In a scenario of iron deprivation, the Fe-S clusters are not formed and Aft1/2 are free to bind to their target genes, involved in several iron transport mechanisms (Outten & Albetel 2013)

Fundamental for the transcriptional iron response network in yeast are iron-sulfur clusters (Li et al. 2012). When iron levels are high, Fe-S clusters are synthesized inside the mitochondria (Fig 1.2-1). These clusters are then transported to the cytosol by the mitochondrial transporter Atm1 (Fig 1.2-2), where they bind to glutathione (GSH), forming [2Fe-2S] bridged complexes with the monothiol glutaredoxins Grx3 and Grx4. These complexes, together with the cytosolic proteins Fra1 and Fra2 travel inside the nucleus and bind to Aft1/2 (Fig. 1.2-3), leading to the dissociation of these transcription factors from their target DNA (Fig. 1.2-4). At the same time the iron signal inside the nucleus leads to Yap5 activation by a yet unknown mechanism (Fig. 1.2-5) and consequently expression of its target genes (Fig. 1.2-6). Among them, CCC1 acts as an iron detoxifier by importing it inside the vacuole and lowering the cytosolic iron pool (Fig. 1.2-7), while Tyw1, an endoplasmic reticulum (ER)-bound protein, sequesters iron as protein-bound Fe-S clusters (Fig. 1.2-8). Yap5 also promotes the expression of Grx4, which maintains the Aft1/2 inactivation and is thought to act also as an iron detoxifier by directly binding to it. On the other hand, when the iron levels are scarce, the Fe-S clusters do not form and Aft1/2 are free to bind to their dependent genes, activating them. The net effect will be the import of iron inside the cell and the shutdown of the mechanisms that keep it inside the vacuole and bound to proteins as Fe-S clusters (Outten & Albetel 2013).

Another layer of iron regulation is present in yeast, which relies on the mRNA-binding proteins (Fig. 1.3). Key to this post-transcriptional control are two mRNA-binding proteins Cth1 and Cth2 (Outten & Albetel 2013). This mechanism can be considered as a fine-tune of the iron metabolism when iron levels are low, as its impact is much smaller when compared to the transcriptional control of Aft1/2 and Yap5. In fact, the expression of Cth1 and Cth2 is directly dependent of the action of Aft1/2. These two proteins transport the respective target mRNAs to the cytosol and promote their degradation, by binding to AU-rich elements (AREs) present in the 3' UTR (untranslated region) (Martínez-Pastor et al. 2013). The proteins under this control are involved in Fe-rich pathways, like respiration, tricarboxylic acid (TCA) cycle, haem biosynthesis, amino acid, sterol and fatty acid metabolism, and mitochondrial Fe-S cluster biogenesis. Interestingly, both Cth1 and Cth2 have AREs in their own 3' UTR, allowing a precise cross and auto-regulation of their expression (Fig 1.3) (Outten & Albetel 2013).

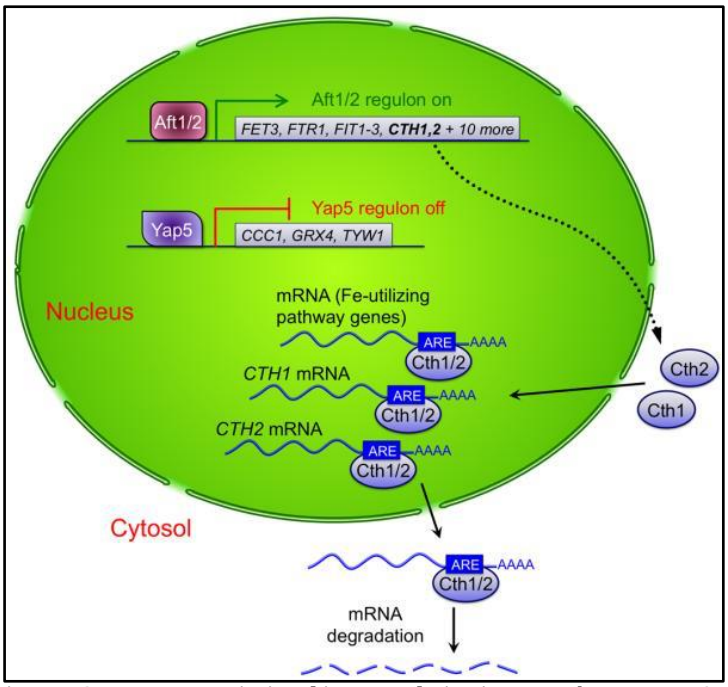


Figure 1.3 - Post-transcriptional iron regulation in yeast: the two mRNA-binding proteins Cth1 and Cth2, whose expression is activated by Aft1/2 upon iron deprivation, bind to AU-rich elements (AREs) in the 3' UTR of the respective mRNAs and transport them to the cytosol where they will be degraded. These genes are involved in several iron utilizing pathways. Additionally, Cth1 and Cth2 themselves have AREs in their own 3' UTR, allowing a precise regulation of their expression. (Outten & Albetel 2013)

1.4 – Iron sensing and regulation in mammalian cells

In mammals, the vast majority of iron is present inside the RBCs as part of haem, the prosthetic group of haemoglobin (De Domenico et al. 2008). Additional iron storages are macrophages and the myoglobin of muscles, while excess iron in stored in the liver (Wang & Pantopoulos 2011). As mammals do not possess mechanisms for iron excretion, the control of iron levels depend on the regulation of dietary uptake in duodenum and macrophage recycling of erythrocytic iron. Iron is absorbed from the intestinal lumen to enterocytes by DMT1 (Divalent Metal Transporter 1) (Gunshin et al. 1997) or as haem-bound by haem transporters, and further from enterocytes to the bloodstream by ferroportin. Macrophages, which recycle iron by consuming senescent RBCs, also have ferroportin as the iron transporter to the bloodstream. Once found in the plasma, iron binds to the glycoprotein transferrin, which serves as a vehicle that transport iron to all the tissues. Transferrin-bound iron is then taken up by receptor-mediated endocytosis upon binding to transferrin receptor 1 and 2 (TfR1 and TfR2). While TfR1 is ubiquitously expressed, TfR2 is predominantly expressed in the liver. Once it is found in the cytosol, iron is transported into the mitochondria to be used for synthesis of haem and Fe-S clusters, while the excess is stored in ferritin, a cytosolic protein. The efflux of iron from enterocytes and macrophages is tightly regulated by the hepatic hormone hepcidin, which upon inflammation or high iron levels targets ferroportin to degradation. On the other hand, hepcidin expression is repressed by iron deficiency and increased erythropoiesis (Hentze et al. 2010).

This systemic regulatory system is coordinated with an intracellular one, that makes use of mRNA-binding proteins. The two proteins involved in this system are iron regulatory proteins 1 and 2 (IRP1 and IRP2). While IRP2 acts only as a mRNA-binding protein, IRP1 also has aconitase activity (Wilkinson & Pantopoulos 2014). Aconitase is an enzyme of the TCA cycle that catalyses the stereo-specific isomerization of citrate to isocitrate via *cis*-aconitate (Beinert & Kennedy 1993). The dual function of this protein is due to the assembly of a [4Fe-4S] cluster, which is absent in IRP2. When iron levels are low, these two proteins act in the respective targets by binding to highly conserved hairpin structures of 25-30 nucleotides, the iron responsive elements (IREs). The major difference between this system and the one present in yeast relies on the localization of the IREs in the target gene's mRNA. Depending if the IREs is located in the 3' or the 5' UTR, the binding of the IRPs will lead to the stabilization or translational blocking of the target mRNA, respectively (Fig. 1.4) (Wilkinson & Pantopoulos 2014). When iron levels are low, both IRPs bind to the 5' UTR of both subunits of ferritin (H-ferritin and L-ferritin), blocking its translation, and to the 3' UTR of transferrin receptor 1, protecting the mRNA from endonucleolytic degradation. The combined effect of blocking iron storage inside ferritin and transport of iron by transferrin upon binding to TfR1 will restore the iron levels inside cells. In the opposite scenario, i.e., iron overload, IRP1 binds Fe-S cluster and acquires aconitase activity and consequently loses its mRNA binding activity, while IRP2 is degraded. The absence of IRP1 and IRP2 allow the translation of ferritin and consequent storage of iron excess, while the mRNA for TfR1 becomes vulnerable for endonucleolytic degradation. The drop in the number of transferrin receptors decreases the cellular iron transport, and consequently restores normal iron levels. More recently, IRE/IRP regulatory network was investigated on the genome-wide level by immunoprecipitation of ribonucleoprotein complexes containing IRP and bound mRNA followed by microarray analyses (Sanchez et al. 2011). In this way, further 35 novel target mRNAs that bind both IRP1 and IRP2 were identified, leading to a much more comprehensive understanding of the IRP regulon in mammalian cells.

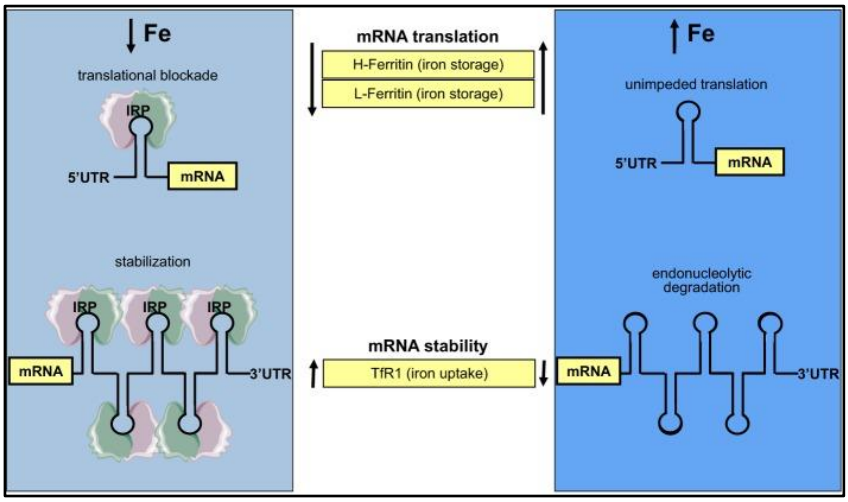


Figure 1.4 - Post-transcriptional regulation by IRPs/IREs in mammals: the mRNA-binding proteins iron regulatory proteins 1 and 2 (IRP1 and IRP2) block or promote their target mRNAs translation by binding to iron responsive elements (IREs) in the 5' or 3' UTR respectively. Upon iron deprivation, they block the translation of both ferritin subunits and stabilize transferrin receptor 1 (TfR1), while upon iron excess they promote the unimpeded translation of the ferritin subunits and promote the endonucleolytic degradation of TfR1 mRNA. Figure from (Wilkinson & Pantopoulos 2014).

1.5 – Iron in *Plasmodium*: what do we know?

For all the reasons described above, it is clear that iron possesses a central role for living organisms, and *Plasmodium* is not an exception. Despite that, it is surprising how little is known concerning the mechanisms employed by the parasite to deal with this resource. Using iron chelating compounds, it has been shown that iron is essential for the survival of the parasite across the life cycle: from gametocytes and blood stage forms (Ferrer et al. 2012) to liver stage forms (Stahel et al. 1988).

So far, two iron transporters have been identified in *Plasmodium*: ZIP domain-containing protein (ZIPCO) (Sahu et al. 2014) and Vacuolar Iron Transporter homologue (VIT) (Slavic et al. 2016). ZIPCO was shown to localise to the plasma membrane of rodent *P. berghei* parasites, and to be predominantly expressed during the liver stage development of parasites. Its gene deletion greatly impairs the parasite's development in hepatocytes, but has no effect on the ability to infect mosquitoes and multiply in mouse blood (Sahu et al. 2014). VIT is a homologue of yeast's CCC1 and plant Vacuolar Iron Transporters that is expressed in *Plasmodium* throughout the life cycle and provides an iron detoxification mechanism, by removing iron excess from cytosol (Slavic et al. 2016). As it happens with ZIPCO, VIT is not essential during the entire life cycle, which indicates the presence of compensatory mechanisms to transport and detoxify iron.

Haem biosynthesis is directly dependent on iron, and has been shown recently to play a role in the mechanism of action of artemisinin, the main drug used to treat the disease. During the intraerythrocytic developmental cycle the parasite degrades host haemoglobin, producing amino acids and releasing free haem, which is sequestered in the crystal hemozoin. The genetic down modulation of falcipains 2 and 3, two cysteine protease haemoglobinases, confers artemisinin resistance in rings, confirming the role of haemoglobin degradation in artemisinin's mode of action (Xie et al. 2016). It seems that haem synthesized by the parasite during the ring stage and resultant from haemoglobin degradation in later stages, rather than free iron, is the main activator of artemisinin (Wang et al. 2015).

It is intriguing that even though degradation of haemoglobin leading to the release of haem plays such a central role in the survival of blood stage parasites, we still do not know what exactly is the source of iron that parasites use during their replication. Most of the release haem is converted into inert crystal hemozoin, preventing toxicity of excess haem. Evidence so far suggests that parasites cannot enzymatically degrade this haem, which would suggest that it could not serve as a source of iron (Sigala et al. 2012). Even less studied and known is the case of parasite iron metabolism at their liver stage of development.

No one can ignore the impact iron has in every aspect of *Plasmodium* biology, from its own survival and replication to all the interactions established with the host. The need for sensing and consequently adapting to iron levels must be of major importance to the parasite, and it wouldn't be a surprise to observe a remodelling of gene expression in response to this resource, as it happens in yeast and mammalian cells. By using one of the most recent and advanced approaches to study genome-wide responses in organisms, carefully described in the next section, this work will try to provide new insights in the complex role of iron for *Plasmodium* parasites.

1.6 – Welcome to the omics' era!

In the past two decades, the development of high-throughput sequencing technologies and bioinformatics tools to analyse the data generated had a major impact in science. The "omics" fields provided different layers for studying cellular and molecular systems: genes (genomics), mRNAs (transcriptomics), proteins (proteomics) and metabolites (metabolomics) can now be profoundly studied. By using one or more of these approaches it is possible to characterize an organism in a highly dynamic way. Genomes from several organisms are already fully sequenced (including *P. falciparum* (Gardner et al. 2002)), and many more will follow in the next few years, providing an invaluable resource for the studies of living organisms. It is also possible to address global responses to specific conditions, for example by using transcriptomics to compare the gene expression between a wild-type and a mutant organism, or proteomics to compare the protein expression between control and treated

condition (administration of a drug or toxic compound, different temperatures or pH levels, etc.). These kinds of studies are starting to become widely used in many areas of research, from evolutionary biology to biomedical sciences, and are of great importance to help solving fundamental questions like the genetic basis of adaptation and the evolution of gene expression profiles, among others.

1.6.1 – Our approach: RNA-seq!

Whole transcriptome shotgun sequencing, a.k.a RNA-seq, is an approach that allows transcriptome profiling using deep-sequencing technologies (Chu & Corey 2012). The transcriptome is commonly defined as the complete set of transcripts and their abundance, in a cell, in a particular stage and condition (Wang et al. 2009). It is becoming widely used due to the advantages compared to DNA microarrays, namely the lower amount of material needed, lower background noise, higher sensitivity, and the fact that it doesn't need prior knowledge about the organism's genome sequence (Wang et al. 2009).

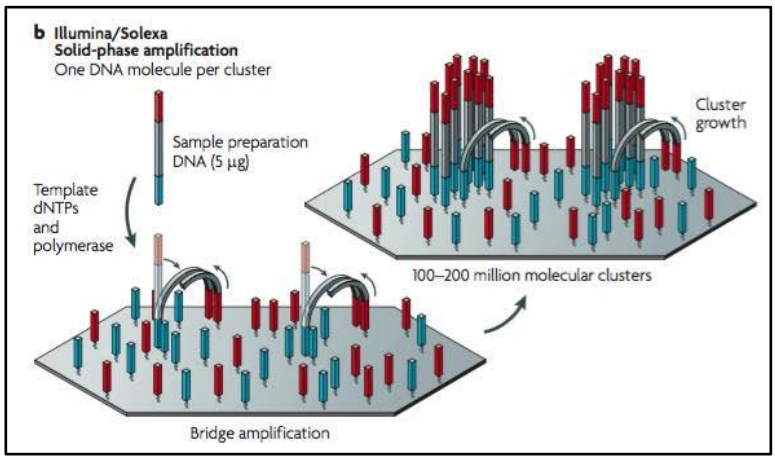


Figure 1.5. – Solid-phase amplification: library is loaded into a flow cell and fragments hybridize to the surface. Each fragment is amplified by bridge PCR into a clonal cluster Figure from (Metzker 2010).

A typical RNA-seq protocols involves the following steps: total RNA isolation, ribosomal RNA depletion (as it represents about 95% of total RNA), cDNA synthesis (library preparation), ligation of adapters to each cDNA molecule to allow the binding to a solid support (cluster generation) and parallel sequencing of millions of locally amplified cDNA fragments via sequencing by synthesis (in this step, details vary according to the platform used) (Hoeijmakers et al. 2013). In our particular case, the platform used was Illumina. After the cDNA synthesis, cDNA molecules are bound to specific primers, which allows the binding to a solid support, where amplification occurs by bridge PCR. This is called solid-phase amplification (Fig. 1.5) (Metzker 2010). After this step, clonally amplified fragments are sequenced by cyclic reversible termination (CRT). In this phase, each nucleotide (A, T, C, G) has a specific fluorophore incorporated, as well as a terminating group. Upon binding of each nucleotide, mediated by a DNA polymerase, the incorporation is terminated and the signal in that position is recorded (Fig 1.6) (Metzker 2010). As

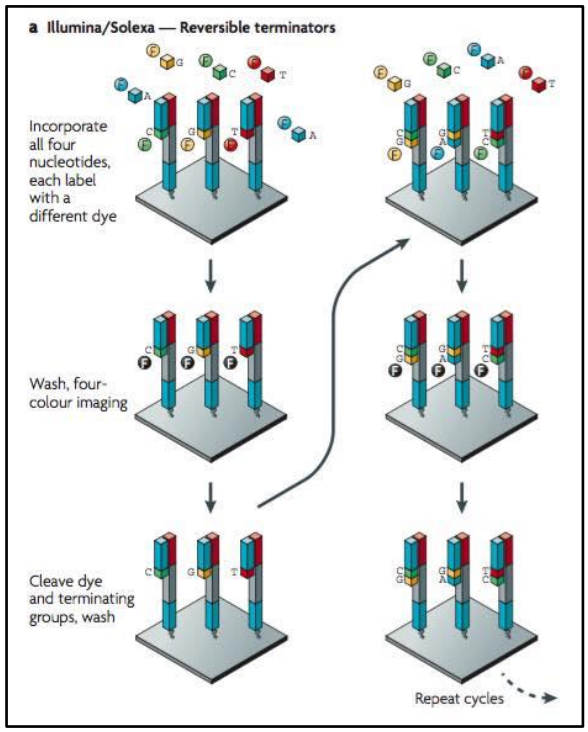


Figure 1.6. – Cyclic reversible termination: fluorescently labelled nucleotides are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded Figure from (Metzker 2010).

the four nucleotides are present during each sequencing cycle, natural competition between them minimizes incorporation bias. Also, the separation between incorporation and recording of the signal increases the precision of the technique.

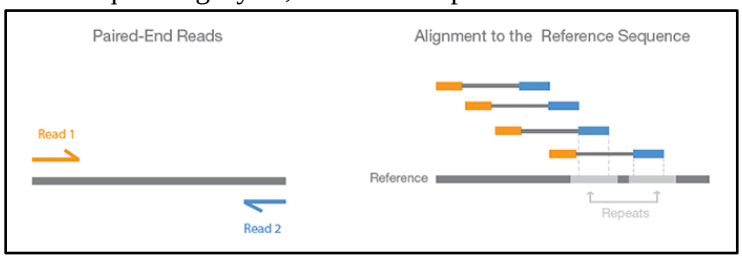


Figure 1.7. – Paired-End Reads: both ends of the cDNA fragment are sequenced, providing a read pair to be aligned to the reference sequence Figure from (Illumina 2015)

Another advance regarding this technology is the use of paired-end reads. This involves sequencing each end of the cDNA fragments in the library and aligning the forward and reverse read as read pairs. Because the distance between each pair is known, alignment algorithms can use this information to map the reads more precisely (Fig. 1.7) (Illumina 2015).

There are several type of questions that might be addressed using RNA-seq, from the detection of alternative splicing events (Griffith et al. 2010), detection of non-protein coding RNAs (Broadbent et al. 2015), detection of fusion genes (Maher et al. 2009), among others. Despite that, the most common and straightforward application of RNA-seq is still the quantification of gene expression between samples, in order to find differentially expressed genes (Oshlack et al. 2010). As referenced above, several variables might be tested as factors affecting gene expression, from treatments with drugs to alterations in environmental conditions, or even compare the transcriptome of a wild-type and a mutant organism of the same species. Also, there are studies addressing transcriptomic changes in response to iron: it is described that yeast remodels its transcriptome in response to iron deprivation, switching from iron-dependent to iron-independent metabolic pathways by the activation of the iron independent transcription factor Aft1p (Shakoury-Elizeh et al. 2004) . In a subsequent paper, also using yeast in an iron deprivation condition, the authors identified 100 and 42 genes up and down-regulated, respectively. These represents around 2% of the genome, and most of them were associated with iron uptake and transport mechanisms (Jo et al. 2009). These works clearly show the potential of these approaches to study iron responses at a genome-wide level. Additionally, in a study from Moyerbrailean et. al, the authors analysed the transcriptomic profile by RNA-seq in response to 23 environmental perturbations (including changes in iron levels) in lymphoblastoid cell lines (Moyerbrailean et al. 2015), showing that these techniques will become common practice in several laboratories, due to the optimization of sequencing protocols and the decrease in cost.

These gene expression profiling studies are targets of great interest for studying pathogen's responses to compounds, in order to find potential drugs. Indeed, this approach was already used in *Plasmodium falciparum*, both using microarrays (Hu et al. 2010; Siwo et al. 2015) and RNA-seq (Shaw et al. 2015). The fact that these parasites are easily maintained in continuous culture (Trager & Jensen 1976), facilitates *in vitro* drug testing.

1.7 – *Plasmodium falciparum* blood stage transcriptome: a unique gene regulation cascade mechanism among eukaryotes

As described in the first section, the parasite goes through a complex sequence of events during the 48h blood stage, assuming three different stages with particular morphological and metabolic features: ring (0-18h), trophozoite (18-30h) and schizont (30-48h). In order to complete this cycle and consequently establish an infection in the human host, it must have a tight temporal regulation of transcription, as

each stage is present within a specific window of time. This has been a fundamental question in the malaria parasite research, as it would deeply improve the knowledge of its biology and might open doors for the development of stage specific drugs or vaccines. Additionally, the gene expression regulatory elements themselves might be good targets for new antimalarial drugs, as the deregulation of the mechanisms that control the developmental cycle would prevent parasite growth and potentially eliminate it. The sequencing of *P. falciparum* genome was a major tool to elucidate the molecular mechanisms behind these developmental events. The 22.8 Mb genome encodes around 5400 genes and is one of the most AT-rich genomes among living organisms. It is organized in 14 nuclear linear chromosomes, a circular apicoplast genome and a linear mitochondrial genome (Gardner et al. 2002). While the nuclear genes' expression is not coregulated, the plastid genes are organized in polycistronic units and highly coregulated (Bozdech et al. 2003).

Indeed, the first study of *P. falciparum* IDC transcriptome revealed amazing results. Using microarrays, Bozdech and Llinás deeply characterized the transcriptome during the entire IDC, by measuring mRNA levels of highly synchronized parasites during the 48 hours at a 1-hour time scale resolution (Bozdech et al. 2003). With these results, they observed that around 80% of the genes expressed have changes in transcript abundance across the IDC. More surprising is the fact that the parasite exhibits a continuous cascade of gene expression, with most of the genes being expressed only once during the IDC, with its maximum expression level correlating with the specific function of the product produced (Fig. 1.8 left) (Bozdech et al. 2003). During the first half of the IDC, equivalent to ring/early troph stages, genes involved in general cellular mechanisms are activated, i.e., transcription, translation, glycolysis and ribonucleotide synthesis (Fig. 1.8 left B-E). With the progression to the trophozoite stage comes the expression of genes involved in several metabolic pathways. This result is expected, as the trophozoite stage is the most metabolically active. The class of genes characteristic of this stage include DNA

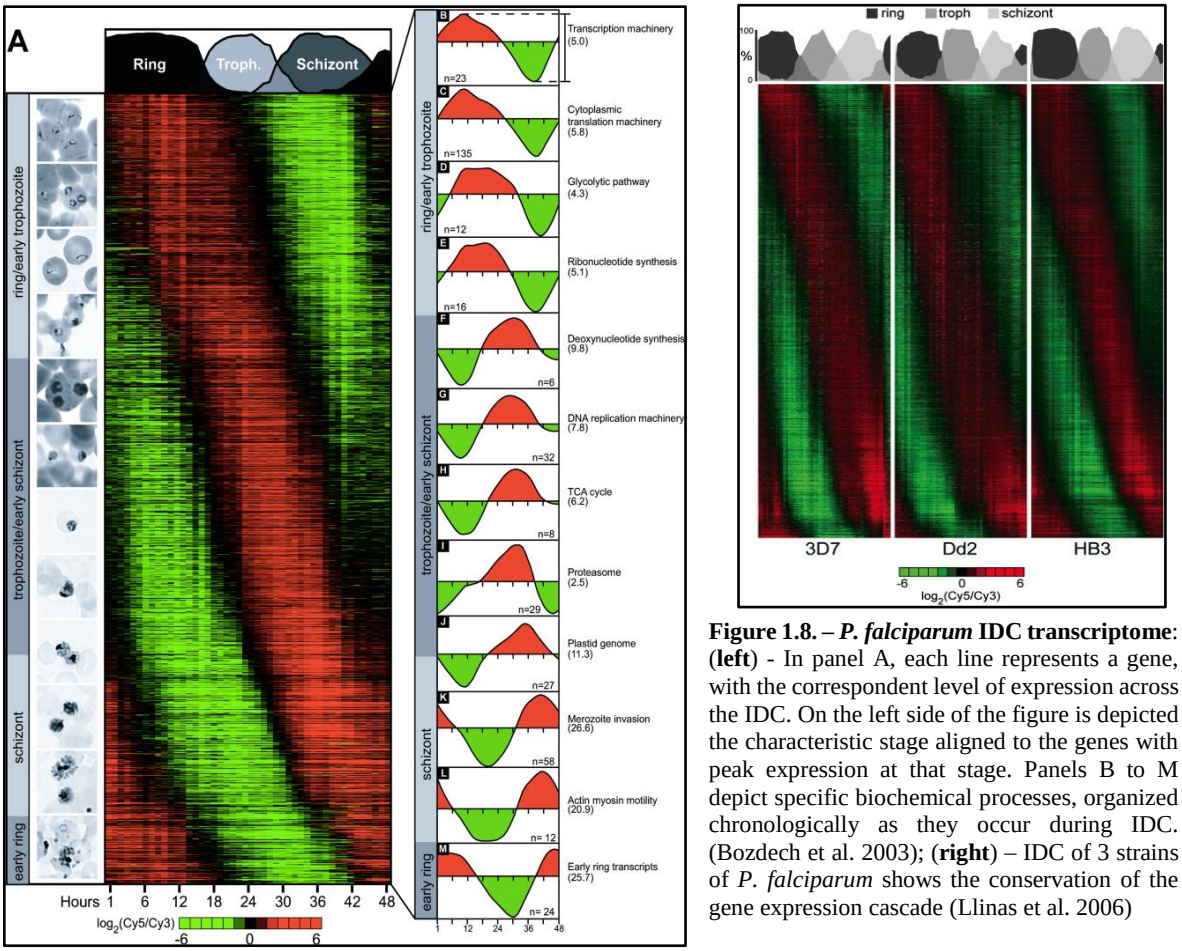


Figure 1.8. – *P. falciparum* IDC transcriptome: (**left**) - In panel A, each line represents a gene, with the correspondent level of expression across the IDC. On the left side of the figure is depicted the characteristic stage aligned to the genes with peak expression at that stage. Panels B to M depict specific biochemical processes, organized chronologically as they occur during IDC. (Bozdech et al. 2003); (**right**) – IDC of 3 strains of *P. falciparum* shows the conservation of the gene expression cascade (Llinas et al. 2006)

replication machinery, TCA cycle and deoxynucleotide synthesis (Fig. 1.8. left F-H), showing the coordinated transcription of genes involved in DNA synthesis and the precursors required for this process. Also, these findings clearly separate deoxynucleotide production as a trophozoite-process and ribonucleotide production as a ring-process. The shift to schizont stage induces genes involved in proteasome and plastid genome (Fig. 1.8. left I-J), indicating a role of protein degradation in the development of the parasite. On the other hand, the activation of plastid genome clearly shows the expression of parasite-specific mechanisms with the progression of IDC. In the final part of IDC, genes involved in host invasion are activated, as new merozoites are produced. These categories include merozoite invasion and actin myosin motility (Fig. 1.8. left K-L). The latter shows the implication of action and myosin in the remodelling of host cells, allowing the successful invasion. With the successful merozoite invasion and consequent formation of new rings, comes the expression of early ring transcripts (Fig. 1.8. left M).

In a subsequent work, Llinás and Bozdech characterize the transcriptome of two additional *P. falciparum* and were able to show that this pattern was conserved between them (Figure 1.8. right), clearly placing it as a specific parasite phenomena (Llinas et al. 2006). Le Roch et al. observed similar results in their study of the transcriptome characterization of nine different stages of the parasite (including asexual IDC stages), like the high percentage of genes that are expressed only once during the IDC, the shift of functions from ring and trophozoite to schizont and the fact that almost half of the genes are cell cycle regulated (K G Le Roch et al. 2004).

The big question imposed by these results is how is this gene expression cascade across the IDC regulated. There are several evidences suggesting post-transcriptional and translation control outweighing transcriptional control: first, the reduced number of genes encoding transcription factors and transcription associated proteins in *P. falciparum* genome (Coulson et al. 2004) (despite the identification of the Apicomplexan AP2 family of DNA binding proteins as potential transcription factors controlling development) (Painter et al. 2012). Second, the significant delay observed between mRNA and protein abundance also imply mechanism of post-transcriptional and translational control (Karine G Le Roch et al. 2004; EM et al. 2014). However, Caro et al. recently observed that transcription and translation are closely coupled for most genes (Caro et al. 2014), giving strength to the debate. Third, several studies suggest that mRNA-binding proteins have a huge impact in gene expression regulation of the parasite: Reddy et al. performed a bioinformatic identification and annotation of 189 putative RNA-binding proteins of the parasite (Reddy et al. 2015). Additionally, in a very recent work, Bunnik et al. made a deep comparative genomics study complemented with experimental validation of the "mRNA-bound proteome", being able to observe that *P. falciparum* has a higher number of genes encoding mRNA-binding proteins when compared to other organisms (Bunnik et al. 2016). Among these, PfAlba1 seems to have an important role in regulating time translation during IDC, as it binds to several mRNAs changing their steady state levels (Vembar et al. 2015). Finally, epigenetic mechanisms seem to play a role in transcriptional control of the parasite: Rovira-Graells et. al observed that the histone mark H3K9me3 (which promotes the formation of heterochromatin in higher eukaryotes) has a function in gene repression in the parasite (Rovira-Graells et al. 2012).

All the findings described above illustrate the importance of deep understanding of the regulation of the transcriptome across the IDC. However, it is also clear that these studies unveiled more questions than answers: what are the exact molecular pathways regulating gene expression during IDC? Does the parasite sense its surrounding environment (human host) in order to "decide" when it should start the development to the subsequent stage, or is this a stochastic event? Does it have any kind of quorum-sensing mechanism, proliferating adequately to the density of parasites in order to survive inside the human host without killing it? Also intriguing is the bridge to the host-pathogen battle for resources: are

there any mechanisms of sensing and consequent gene expression regulation in response to the host nutritional status? This topic will be particularly addressed in this thesis, making use of the tools provided by the boost of high-throughput sequencing with special emphasis in iron. Does the parasite have mechanisms of transcriptional and post-transcriptional regulation of gene expression in response to iron like yeast and mammalian cells? If it does, which genes are target of this regulation: iron transporters, detoxifiers or Fe-S cluster proteins? We hope to be able to start answering these (and many more) questions, which are of great relevance for the better understanding of the basic biology of the parasite, allowing for the development of broader (in terms of activity across the life cycle stage) and at the same time, more specific (in terms of targeting only the parasite and not the host) anti-malarial strategies.

2 – Thesis aims

As iron is a micronutrient with a major role in infection, being essential for both the pathogen and the host, the main objective of this work was to identify genes involved in iron sensing and regulation in *Plasmodium falciparum*, the parasite responsible for the most severe form of Malaria in humans. To address that, putative homologues of iron-responsive genes in model organisms were identified in *P. falciparum* and it was tested how their expression levels were affected by changes in the iron concentration of the parasite environment. However, since no homologues of the main iron-sensing proteins could be identified in *Plasmodium*, an RNA-seq experiment was performed in order to identify transcriptomic changes in *P. falciparum* exposed to iron excess and iron depletion condition. In this way, by identifying differentially expressed genes between the conditions tested through a genome-wide approach, we intend to identify new mechanisms of iron-sensing and regulation of the complex iron metabolic network in the malaria parasite.

3 – Methods

3.1 BLAST (Basic Local Alignment Search Tool) and Multiple Alignment

BLAST searches were performed in the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the protein blast (blastp) tool with default parameters. After the BLAST search, multiple alignments were performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters between the potential homologue in *P. falciparum* and the one used in the BLAST search. All protein sequences were provided in FASTA format (retrieved from Uniprot or PlasmoDB for *S. cerevisae* and *S. pombe* or *P. falciparum* genes, respectively). In order to quantify the gene expression by qPCR, primers were designed using the PrimerQuest tool in the IDT website (http://eu.idtdna.com/).

3.2 – Preparation of biological samples

3.2.1 – *P. falciparum in vitro* culture

P. falciparum 3D7 strain blood stage parasites were cultured using standard protocol for *in vitro* culture (Trager & Jensen 1976). The parasites were kept in a culture flask in malaria complete medium (MCM) and blood diluted to 5% haematocrit. The incubator's temperature was 37°C, the O_2 and CO_2 percentage 4 and 5, respectively.

Medium was changed every day or every two days, depending on the parasitemia (percentage of infected RBCs). Smears were made regularly to keep track of parasitemia and dilute the culture with MCM and blood if needed. Levels of parasitemia higher than 10% might lead to culture's crash and consequent death of the parasites.

3.2.2 – Giemsa staining of smears

An aliquot of *P. falciparum* culture was transferred to a microcentrifuge tube and centrifuged at 2000 rpm for 1 min. Then, 2 μl of pelleted RBCs were aspirated and smeared on a microscope slide. Smears were dried for 1-2 min, fixed in methanol for 10 seconds and stained in Giemsa Stain modified solution (Sigma) diluted 5% in Giemsa buffer for 5-10 minutes. Parasitemia was counted in the optical microscope under the 100x objective (with the use of immersion oil).

3.2.3 – MCM composition

- 500 ml RPMI Medium 1640 (1X) [-] L-Glutamine (gibco® by Life Technologies)
- 12.5 ml HEPES Buffer Solution (1M) (gibco® by Life Technologies)
- 500 μl Gentamicin 50 mg/ml (gibco® by Life Technologies)
- 5 ml L-glutamine 200mM (100X) (gibco® by Life Technologies)
- 50 ml Albumax II solution (preparation below)

3.2.4 – Albumax II solution composition

- 5.2 g RPMI 1640 powder (gibco® by Life Technologies)
- 500 μl Gentamicin 50 mg/ml (gibco® by Life Technologies)
- 2.98g HEPES (gibco® by Life Technologies)
- 1g Glucose
- 1.67g $NaHCO_3$
- 0.1 g Hypoxanthine
- 25g Albumax II (gibco® by Life Technologies)
- Distilled water was added up to 500ml, and the solution was filter sterilized

3.2.5 – Washing of Human Blood

Total human blood was centrifuged at 2000 rpm for 5 min and the supernatant was removed in order to deplete serum and leukocytes. Then, pelleted blood was washed in PBS 1x up to 50 ml for two times. Washed blood was centrifuged and ressuspended in MCM at a ratio 1:1 and kept at 4°C.

3.2.6 – Thawing of glycerolyte-frozen parasites

A frozen vial of *P. falciparum* 3D7 was removed from cold storage and thawed at 37°C for 1-2 min. After thawing, blood was transferred to a 50 ml falcon tube with a sterile pipette and blood volume (V) was measured. Then, 0.1 x V of 12% NaCl in distilled water was added to the blood slowly, dropwise, while shaking the tube gently and the tube was let to stand for 5 minutes. Then, 10 x V of 1.6% NaCl in distilled water was added to the blood slowly, dropwise, swirling the tube and blood was centrifuged at 500 g at 20°C for 5 min. Supernatant was aspirated and 10 x V of MCM was added slowly, dropwise, while shaking the tube. Blood was centrifuged at 1500 rpm at 20°C for 5 min and supernatant was aspirated. Pelleted blood cells were ressuspended in MCM with 5% haematocrit and transferred to a culture flask. Smears were made and medium was changed every day until the parasites become visible by Giemsa staining, which happened around day 3.

3.2.7 – Sorbitol synchronization

Parasites should be taken when they are mostly at ring stage (they must not be later than 10 to 12h post-invasion when the sorbitol treatment is done). Culture was transferred to a falcon tube and centrifuged at 600 g for 5 min. Then, supernatant was removed and 10 ml of sorbitol 5% in distilled water was added. Tube was left for 10 min at room temperature, shaken 2 or 3 times and centrifuged at 600 g for 5 min. Blood pellet was washed in PBS, centrifugation was repeated and blood was washed in MCM. Finally, blood was centrifuged and ressuspended in MCM diluted to 5% haematocrit (as previously described). In order to keep parasites synchronized, the procedure was repeated after 1 cycle (approximately 48h).

3.2.8 – Iron Treatments

Iron excess conditions was achieved by adding $FeSO_4$ (100 mM stock in 0.1M HCl) to the medium to a final concentration of 200 μM. Note: in order to keep Fe in the 2^+ form (the one permeable to biological membranes) ascorbic acid (Aa) solution (stock 100 mM) was also added to a final concentration of 200 μM. In order to be able to make valid comparison, a condition with only ascorbic acid added to the medium was always present. Iron deficiency condition was achieved by adding deferoxamine (DFO) to the medium (stock 50 mm) to a final concentration of 20 μM. To simplify the nomenclature, iron excess treatment will be referred as Fe, while iron depletion will be referred as DFO. Each condition had a correspondent control: for Fe it was a condition with ascorbic acid added to the medium (referred as Aa) and for DFO it was a control with normal MCM without any treatment (referred as Control).

3.2.9 – PhenGreen assay

In order to make sure that the treatments were changing the iron levels, the labile iron pool of the iRBCs was measured by flow cytometry using the PhenGreen fluorescent dye, after 6 and after 24 h incubation (Note: all the samples were processed in parallel, so the treatments were set up accordingly, i.e., 24 hour treatments were first set up and after 18 hours 6 hour treatments were set up. In this way, both treatments finished at the same time).

In a plate, each treatment was set up in duplicate in order to analyse both time-points as described above. The parasites were mostly in the trophozoite stage (sorbitol synchronized) with a parasitemia of 1-2% by the time of the collection for the analysis. When the incubation time was finished, the plate was centrifuged at 500 g for 5 min and medium was removed. Syto61 staining solution was prepared for DNA staining as well as PhenGreen solution in RPMI medium without FBS at 0.5μM and 10μM respectively. This step is of major importance as FBS might quench fluorescence. After 1 hour incubation, samples were washed 2 times in PBS. Finally, stained cells were analysed in a BD Accuri C6 flow cytometer, where PhenGreen-positive cells were detected in FL1-A channel and Syto61-positive cells in FL4-A channel. The geometric mean of PhenGreen fluorescence for the FL1-A, FL4-A positive subset was determined for all samples.

3.2.10 – Viability assay

Each treatment was set up (Fe + Aa, Aa and DFO) as well as a control situation without any treatment in a plate (as described above). The parasites were mostly in trophozoite stage (sorbitol synchronized)

with a parasitemia of 1-2%. After 6 hour smears were made, then the plate was centrifuged at 500 g for 5 min and the medium removed. Cultures were replaced with MCM without any treatment, and after 24 hour smears were repeated.

3.2.11 – RNA extraction

Before starting, centrifuge was cooled down at 4°C for the Tryzol phase separation. Gloves were always used and RNAse free conditions were followed, namely tips, solutions, etc. Until the first wash work was performed under the chemical hood and waste discarded into halogenated residues due to the use of chloroform.

Cultures were transferred to a falcon tube and centrifuged at 500 g for 5 min. The total volume of the blood pellet was calculated for each sample and Tryzol LS was added at a 1:3 ratio (for example, for 200 µl of blood 600 µl of Tryzol LS were added). These mixtures were quickly ressuspended and sat at room temperature for 5 min in order to allow complete RBCs lysis. Then, 240 µl of chloroform were added for 1 ml of Tryzol LS added. Samples were shaken vigorously by vortex and incubated at room temperature for 5 min. Following incubation, samples were centrifuged at 12000 g for 25 min at 4°C. While centrifugation was running, aqueous phase collection tubes, columns and elution collection tubes were labelled and DNAse mix was prepared. When centrifugation was done, upper aqueous phase was transferred (avoiding white interphase which contains DNA) to a new labelled tube and volume was measured. Afterwards, 1.5x volume of 100% ethanol were added followed by pipetting up and down. Approximately 700 µl was loaded to a labelled NZY binding column (blue ring), centrifuged at 13000 g for 30 seconds at room temperature and flow-through was discarded. Any remaining sample was loaded to the column and centrifugation was repeated. Then, following the NZY kit protocol, for each sample DNAse mix was prepared by adding 10 µl of DNAse I into 90 µl of digestion buffer. 95 µl of the DNAse mix was applied into the middle of the column and columns were incubated for 15 min at room temperature. Proceeding to washes, samples were washed once with 200 µl of NWR1 and centrifuged at 11000 g for 30 seconds at room temperature. Following the first wash, samples were washed 3 times with 500 µl of NWR2 supplemented with 100% ethanol and centrifuged at 11000 g for 30 seconds at room temperature. Meanwhile, 50 µl of RNAse free distilled water per sample were transferred to an Eppendorf tube and heated to 56°C, which were used for elution of RNA. Columns were transferred to new collection tubes and centrifuged at 13000g for 2 min at room temperature. Then, tubes were left open for 1 min to allow evaporation of ethanol, phenol, etc. Afterwards, columns were transferred to new RNAse free 1.5 ml tubes and eluted carefully by adding 50 µl of pre-heated H_2O to the middle of the column. Lids were closed and water was incubating for 2 min to allow H_2O to absorb. Finally, samples were centrifuged at 11000 g for 1 min at room temperature, RNA concentration was measured in Nanodrop and stored at -80°C.

3.2.12 – cDNA synthesis

For each sample, the volume needed to have 400 ng of RNA was calculated and RNAse free H_2O was added up to 20 µl.

cDNA mix was prepared using NZYTech reagents for a final volume of 20 µl:

- 4 µl 10x reaction buffer
- 2 µl dNTPs NZY mix (10 mM each)

- 2 µl Random hexamer mix 50 ng/µl
- 1 µl NZY Ribonuclease inhibitor
- 2 µl NZY reverse transcriptase
- 9 µl RNAse free H_2O

20 µl of mix and 20 µl of RNA from each sample were pipetted in PCR tubes from Qiagen.

Tubes were placed in a T100 Thermal Cycler from Bio-Rad and the following cDNA synthesis protocol was used:

- 25°C for 10 min
- 50°C for 50 min
- 85°C for 5 min

Finally, tubes were stored at 4°C or -20°C for short or long-term storage, respectively.

3.2.13 – Quantitative PCR (qPCR)

Gene expression was measured by qPCRs using SybrGreen reagents in either 96 or 384-well plate. The mix's composition was the following:

- 0.2 µl reverse primer of target gene
- 0.2 µl forward primer of target gene
- 5 µl SybrGreen mix
- 3.6 µl H_2O

Note: this is the mix for 384-well plate, for 96-well plate twice the volume of all reagents was added.

The mix for each gene was pipetted in duplicate per sample plus one negative control with water instead of cDNA in a RT-PCR plate. Then, 1 µl of cDNA was added directly to each well and 1 µl of water was added to the negative control well. The plate was centrifuged at 3000 rpm for 3 min.

Gene expression was measured in a RT-PCR ViiA 7 System or RT-PCR 7500Fast for 384 or 96-well plate respectively, and was normalized to the average of expression of the house-keeping genes. ΔΔCt were calculated for each treatment and respective control (DFO vs. Control and Fe vs. Ascorbic) and gene expression was obtained by the formula $2^{-\Delta\Delta Ct}$.

3.3 – Sequencing

Prior to the sequencing itself, a quality control was performed for all samples by measuring the RNA integrity number (RIN) and concentration in order to assess the RNA quality and quantity. All samples were within suitable parameters (amount of RNA > 10 µg and RIN > 6.9). Then, the whole-transcriptome library was built using the TruSeq Stranded mRNA Library Preparation Kit. During this procedure, it was necessary to deplete rRNA as it represents 95-99% of total RNA, in order to allow sufficient sequencing-depth for mRNA. The generated DNA fragments (DNA libraries) were sequenced in the Illumina Hiseq 2500 platform, using 100bp paired-end sequencing reads. The raw sequence data underwent automatic initial treatment, namely adapters and low quality bases trimming. All these procedures were performed in the company ®STABVIDA Lda (project reference: NGS2016.0404.01(P))

3.4 – RNA-seq bioinformatic analysis

Note: all the code is written in Menlo font size 9. TopHat and Cufflinks package software are operated through the UNIX shell without graphical user interface. All the softwares are installed in IMM server Biocomp and were run in the server. This protocol is an adaptation of the one used in Trapnell et al. (Trapnell et al. 2012)

The steps until the differential expression analysis are depicted in the Methods section, as they are a common part of the process of studying gene expression using RNA-seq.

3.4.1 – Quality control

All the samples' quality was evaluated using FastQC High Throughput Sequence QC Report version 0.11.4.

3.4.2 – Mapping to reference genome

The mapping of the reads from each sample to the *P. falciparum* 3D7 reference genome was performed using TopHat version 2.0.13, a software that aligns reads to the genome and discover transcript splicing sites. Bowtie is the short read aligner used by TopHat as the alignment "engine". Bowtie is one of the most used and data efficient short read alignment software, as it uses a highly economic data structure called FM index to store the reference genome. This structure allows the reference genome to be searched in a fast and efficient way. Then, TopHat breaks up the reads Bowtie cannot align in smaller pieces called segments, which will be aligned to the genome independently (Trapnell et al. 2012). This version of TopHat uses Bowtie 2.

Before running TopHat, it is needed to build a Bowtie index for the organism used in the RNA-seq experiment. In order to build the index, the reference genome was downloaded in FASTA format from genedb website: http://www.genedb.org/Homepage/Pfalciparum.

- Code to create bowtie index for the *P. falciparum* reference genome (needs prior download of reference genome in FASTA format):

```
bowtie2-build P.falciparum_index.fa Palciparum_index
```

- Map reads for each sample to reference genome using TopHat:

```
tophat /home/FM/msantos/Indexes/Pfalciparum_index /home/FM/msantos/Reads/Sample1-C1-NGS2016040401P094_S1_L001_R1_001.fastq,/home/FM/msantos/Reads/Sample1-C1-NGS2016040401P094_S1_L001_R2_001.fastq
```

Note: the first argument is the path to the reference genome index, while the second argument is the path to both paired-end reads for each sample (left read and right read separated by comma, respectively). While the index is the same for all samples, the path to the corresponding sample should be changed.

3.4.3 – Assembling of transcripts

Cufflinks is a software that assembles individual transcripts from RNA-seq reads that have been aligned to the genome. This is a crucial step for correct gene expression quantification. As samples may contain reads from multiple splice variants of a given gene, Cufflinks makes a parsimonious transcriptome assembly of the data. This means that it tries to infer the splicing structure of each gene and report the smaller number possible of full-length transcript fragments (called transfrags) needed to "explain" the splicing events present in the data (Trapnell et al. 2012). The cufflinks version used was 2.2.1.

- Assemble transcripts for each sample:

```
cufflinks /home/FM/msantos/TopHat_out/Sample_1/tophat_out/accepted_hits.bam
```

Note: the only argument is the path to the file accepted_hits.bam, one of the output files from TopHat. This file is a list of read alignments in BAM format, a binary version of a SAM file. A SAM file (.sam) is a tab-delimited text file that contains sequence alignment data. These formats are described on the SAM Tools website: http://samtools.sourceforge.net.

When working with several RNA-seq reads, it becomes necessary to pool and assemble all of them together in a comprehensive set of transcripts before starting the differential expression analysis. Cuffmerge is a script that makes part of Cufflinks software, and is able to merge several Cufflinks assemblies into a single unified transcript catalogue. Additionally, one can provide a reference GTF file (reference annotation, described in ensemble website: www.ensembl.org/info/website/upload/gff.html), and the script will use it to attach gene names and other metadata to the merged catalogue. The gtf file for *P. falciparum* was downloaded from sanger server ftp://ftp.sanger.ac.uk/pub/project/pathogens/gff3/CURRENT. This file was in gff3 format, so it had to be converted to gtf file (the one accepted by Cufflinks).

- Convert gff3 file to gtf:

```
gffread P.falciparum.gff3 –T –o P.falciparum.gtf
```

Note: the –T option tells the command to convert the file to gtf format, while the –o option tells the command to give the name provided to the output file.

- Create a file called assemblies.txt that lists the path to the assembly file from Cufflinks for each sample, as following:

```
/home/FM/msantos/TopHat_out/Sample_1/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_2/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_3/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_4/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_5/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_6/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_7/transcripts.gtf
/home/FM/msantos/TopHat_out/Sample_8/transcripts.gtf
```

- Run Cuffmerge in all assemblies to create a single reference transcriptome annotation:

```
cuffmerge –g P.falciparum.gtf –s /home/FM/msantos/Indexes/Pfalciparum_index.fa assemblies.txt
```

Note: the –g option allows the user to provide the reference annotation file (in gtf format), while the –s option allows the user to provide the genomic DNA sequences for reference in FASTA format (same file as the one used by TopHat to map the reads). The final argument is the assemblies.txt file with all the path to all the assemblies generated by Cufflinks.

3.4.4 – Differential expression

The program of the Cufflinks package used for differential expression analysis is called Cuffdiff. It is able to calculate the gene expression between two or more conditions and test the statistical significance of the differences between them. The statistical model used assumes that the number of reads that mapped to a gene is proportional to the expression of the same gene, with technical and biological variability being considered as fluctuation factors (Trapnell et al. 2012). Thus, providing biological replicates is essential in order to "teach" the model how the read counts vary between replicates and correctly estimate the significance of the changes between samples.

- Use Cuffdiff with the merged transcriptome assembly from Cuffmerge to test the significance of changes in gene expression levels between conditions:

```
cuffdiff –o CONTROLvsDFO –b /home/FM/msantos/Indexes/Pfalciparum_index.fa –u merged_asm/merged.gtf –L
Control,DFO /home/FM/msantos/TopHat_out/Sample_1/tophat_out/accepted_hits.bam,
/home/FM/msantos/TopHat_out/Sample_5/tophat_out/accepted_hits.bam
/home/FM/msantos/TopHat_out/Sample_4/tophat_out/accepted_hits.bam,
/home/FM/msantos/TopHat_out/Sample_8/tophat_out/accepted_hits.bam
```

Note: the parameter –o sets the name of the directory in which Cuffdiff will write its output, the parameter –b allows the user to provide the genomic sequence the reads were mapped to, in order to detect bias and improve the accuracy of transcript abundance estimates. The parameter –u allows the user to provide the merged transcriptome assembly to more accurately weight read mappings to multiple locations in the genome, while the –L parameter specify a label for each sample which will be included in the output files. The last arguments are the bam files from TopHat for each sample, with replicates being separated by comma (in this case, sample 5 is a replicate of sample 1 and sample 8 is a replicate of sample 4).

Cuffdiff produces several outputs, being the most important a table with the information from the differential expression analysis at the gene level between samples (gene_exp.diff). In this file is present the result of testing the differences in the summed FPKM (Fragments Per Kilobase of transcript per Million mapped fragments) of transcripts sharing each gene_id. The number of reads produced from a transcript is directly proportional to the expression of that transcript in the sample. However, longer transcripts will produce more sequencing reads, because cDNA fragments usually are size-selected during the library construction in order to optimize the output. As such, if gene A is expressed at the same rate as gene B but has twice the size, on average twice as many reads will be generated from gene A compared to gene B. In order to correctly compare the expression levels between two genes, the software must count the reads that mapped to each transcript and normalize the count by each transcript's length. Also, as each run in the sequencing machine might produce a different number of reads, the counts must be normalized for the total yield of the machine. Both these normalization steps are incorporated in the FPKM values (Trapnell et al. 2012).

Additionally, this file has several information from each transcript, like the gene id and the genomic coordinates, but the most important columns are the FPKM in each sample tested, the $\log_2$fold(FPKM_1/FPKM_2) which represents the $\log_2$(fold change) and the p-value.

3.5.5 – Downstream analysis

To visualize sample similarities on a transcriptome-wide scale, we first constructed a distance matrix of expression profiles as follows (adapted from (Broadbent et al. 2015)): the FPKM values were transformed as $\log_2(FPKM+1)$ and the Pearson correlation (ρ) matrix between sample expression profiles was computed using the R function cor(data, method = 'pearson') where data is a table with the FPKM values from each gene in each sample. Then, this matrix was transformed as $(1-\rho)/2$ and MDS (Multidimensional scaling) was performed using the R function cmdscale(matrix_pearson) where pearson_matrix is the matrix from the command cor after transformation. In order to visualize the plot, we plotted each dimension using the R function plot.

To visualize the expression of each transcript across samples and cluster the samples based on these values, we generated a heatmap using the R function heatmap.2 from the gplots package. The venn diagrams were done in R using the function draw.pairwise.venn from the VennDiagram package.

4 – Results

4.1 – BLASTing: in search for putative iron-responsive genes

The first step of the project was to perform a deep search in the literature in order to identify candidate genes whose expression levels might be affected by fluctuations in iron levels. These candidates were chosen from model organisms, namely yeast and mammalian cells (human). As described in the Introduction, iron metabolism and regulation is well studied in these organisms and here we aimed to investigate if genes that are described to alter their expression in different iron conditions are also iron-responsive genes in *Plasmodium*. To that end, protein BLAST searches were performed with the objective of finding homologues in *P. falciparum* (see Methods). For the cases where a potential homologue was identified, a multiple alignment using Clustal Omega was performed to investigate the overall sequence similarity and the regions of the protein that shared the highest sequence identity (see Methods). The E-value obtained through BLAST search is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size. E-value of 1 assigned to a hit signifies that in a database of the current size one might find another sequence with the same score by chance. Therefore, a BLAST result is more reliable when its E-value is closer to zero.

While it was not possible to identify any *P. falciparum* homologues of the main iron sensors of *S. cerevisiae*, Aft1/2 or Yap5 transcription factors, several genes were identified that shared low sequence identity with the two mRNA-binding proteins Cth1 and Cth2. As described in the Introduction, these proteins are expressed by the action of the transcription factors Aft1/2 in conditions of low iron, and act by binding to AREs in the target mRNAs and transport them to the cytosol where they will be degraded. While using Cth1 as a query we found 3 blast hits, and Cth2 only one. For Cth1, the candidate chosen to test further was the one that exhibited higher sequence identity (42%) with a query coverage, i.e, percentage of the query sequence that overlaps the subject sequence, of 19%. For Cth2, the only candidate had 43% of sequence identity. Two issues with this candidate is that the E-value was quite high (1.1) and the query coverage was low (11%). The low query coverage of both these candidates might be explained by the fact that both *P. falciparum* Cth1and Cth2 have a longer amino acid sequence compared to the yeast homologue sequences (843 and 1722 compared to 325 and 285, respectively), and that the regions sharing sequence similarity are the ones involved in mRNA-binding, with the rest of the sequence being significantly different.

S. cerevisiae Grx4, the monothiol glutaredoxin (UniProt ID P32642) involved in the formation of Fe-S cluster and transport of the iron signal to the nucleus in yeast (see Introduction) did not have a hit in *P. falciparum.* As we consider that this protein has an important role in iron metabolism, we performed a BLAST search using the *Schizosaccharomyces pombe* (fission yeast) Grx4 (UniProt ID O74790) and were able to find a hit, PF3D7_0606900, a gene currently annotated in PlasmoDB (Plasmodium Genomics Resource) as a glutaredoxin-like protein (GLP2). *S. pombe* is widely used to study iron metabolism and regulation, and its Grx4 has a similar role as the one from *S. cerevisiae,* being an inhibitory partner of Fep1 under iron depletion conditions, a transcription factor regulating the transcription of several genes involved in iron metabolism (Labbé et al. 2013).

The ER-bound protein Tyw1, which is upregulated upon Yap5 activation in iron excess condition, detoxifies iron excess by binding it as Fe-S clusters. BLAST search identified one putative homologue for Tyw1 in *P. falciparum.* Also, homologues were identified for Tsa1 and Prx1, a molecular chaperone

that might be important for inhibition of Aft1/2 and a mitochondrial peroxiredoxin that might be involved in low iron signaling, respectively (Outten & Albetel 2013). Finally, Mrs3 a mitochondrial iron transporter whose deletion affects the activity of the vacuolar iron transporter Ccc1 in yeast (Li et al. 2010), had two putative homologues in *P. falciparum*, which were named here as MrsA and MrsB.

Our analyses also included homologues of ferrochelatase, the last enzyme of haem biosynthesis pathway located in the mitochondria (Nagaraj et al. 2009), and VIT, the homologue of yeast CCC1 (see Introduction, *Yeast iron sensing and regulation*). Several homologues of human transporters were also tested, namely DMT, ZIP1, ZIP2 (ZIPCO), ZIP3 and CuTP, a copper transporter (Kenthirapalan et al. 2014). In addition, a homologue of mammalian IRP was also included in the analyses, (Loyevsky et al. 2003). Table 4.1 shows all the genes which expression was tested in response to iron, including the house-keeping genes used for normalization.

Table 4.1 – Candidate genes for which expression levels were tested after the iron treatments: the first column describes the gene's short name, the second shows PlasmoDB ID, the third and fourth columns show the BLAST E-value and the percentage of sequence identity determined using Clustal Omega, respectively, and in the last the function (putative if it is not known in *Plasmodium*). Notes: *BLAST search against human homologue **last four genes are the house-keeping genes used to normalize the gene expression levels.

Candidates	*P. falciparum* ID	BLAST E-value	% id (clustal)	Function (some of them only predicted by homology)
Cth1	PF3D7_0906600	0.004	25,3	mRNA-binding protein
Cth2	PF3D7_1134600	1.3	23,16	mRNA-binding protein
Grx4	PF3D7_0606900	3E-22	27,4	Monothiol glutaredoxin
Tyw1	PF3D7_0524900	1E-127	35	Fe-S cluster binding protein
Tsa1	PF3D7_1438900	9E-50	43,01	Molecular chaperone
Prx1	PF3D7_0802200	1E-49	41,98	Mitochondrial peroxiredoxin
MrsA	PF3D7_1241600	1E-15	28,4	Mitochondrial iron transporter
MrsB	PF3D7_0905200	4E-12	31,14	Mitochondrial iron transporter
Ferrochelatase	PF3D7_1364900	3E-15	22,78	Last enzyme of haem biosynthesis pathway
VIT	PF3D7_1223700	6E-8	21,71	Vacuolar Iron Transporter
DMT*	PF3D7_0523800	1E-39	26,92	Divalent Metal Transporter
ZIP1*	PF3D7_0609100	2E-11	24,23	Zn^{2+} or Fe^{2+} permease
ZIP2*	PF3D7_1022300	0,009	18,89	ZIP domain-containing protein (ZIPCO)
ZIP3*	PF3D7_0715900	9E-25	30,52	Zinc transporter putative
CuTP*	PF3D7_0904900	6E-35	23,3	Copper transporter
IRP*	PF3D7_1342100	0	53,74	Iron-regulatory protein
18S rRNA**	PF3D7_1148600	-	-	House keeping
Serine tRNA ligase**	PF3D7_0717700	-	-	House keeping
Inositol 5-phosphatase**	PF3D7_0802500	-	-	House keeping
Adenylosuccinate lyase (ASL)**	PF3D7_0206700	-	-	House keeping

4.1.1 – Iron treatments are changing iron levels in the iRBCs

After having the candidate genes chosen, we developed a strategy to test if their expression is altered in response to iron. For that, we simulated artificial iron depletion and iron excess conditions in blood stage *P. falciparum* parasites *in vitro* (see Methods) to obtain samples for RNA extraction and gene expression analysis by qPCR.

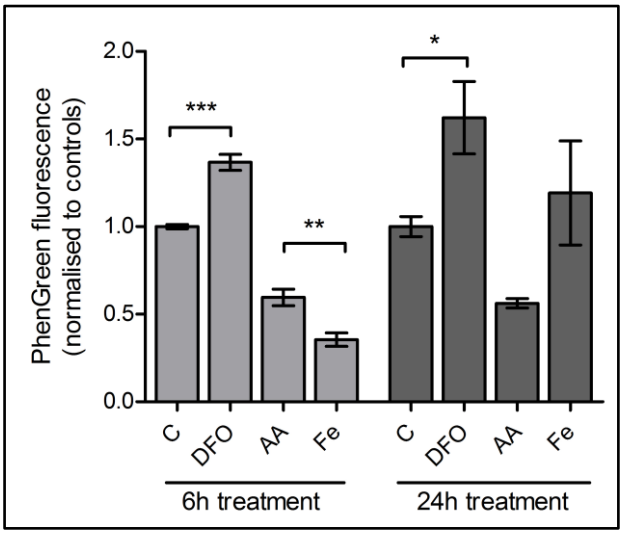


Figure 4.1 – **PhenGreen fluorescence of iRBCs after 6h (left) and after 24h (right) treatments**: C (Control), DFO (deferoxamine 20μM), AA (Ascorbic Acid 200μM), Fe ($FeSO_4$ 200μM + AA 200μM); statistical test: unpaired, two-tailed t-test. Shown are mean + SEM of 4 replicates from 2 independent experiments. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Before starting testing our candidate genes expression in response to iron, we asked whether the treatments were changing the iron levels of the parasite environment. To assess that, we measured the labile iron pool of infected RBCs using PhenGreen (see Methods), a fluorescent dye whose fluorescence is quenched by iron, therefore its fluorescence signal is inversely correlated with iron concentration (more iron results in lower fluorescence and vice-versa). As the fluorescence signal of the controls was different between experiments, we normalized all the fluorescence values to the controls (which were also normalized to the mean of the controls) before pooling the data from independent experiments.

Figure 4.1 clearly shows the effect of the treatments on the labile iron pool of the infected RBCs. After 6 hours, PhenGreen fluorescence of infected RBCs in the DFO condition is significantly higher compared to the control condition, meaning that DFO is efficiently chelating iron. This effect is maintained after 24-hour treatment. In the opposite scenario, after 6 hours of $FeSO_4$ treatment, the PhenGreen fluorescence is lower compared to its control condition (ascorbic acid treatment), showing that with addition of $FeSO_4$ to the medium in the presence of ascorbic acid, the labile iron pool of infected RBCs is higher. However, this effect seems to be lost after 24-hours of $FeSO_4$ treatment, meaning that the parasite might be metabolizing the iron excess or transporting it outside the infected RBCs. It is curious to note that in the AA condition the fluorescence is slightly decreased compared to the control condition, which indicates that AA might be quenching the fluorescence by some iron-independent mechanism. This must be taken into account, as ascorbic acid is also added to the Fe condition in order to keep it in the ferrous form (the iron form permeable to biological membranes, see Methods). Nevertheless, the difference between Fe + AA condition and AA alone after 6 hours is significant.

These results validate our experimental approach to introduce changes in iron concentration in *P. falciparum* infected RBCs. While chelation of iron with DFO is appropriate for both short and long term treatments, by addition of $FeSO_4$ labile iron pool in iRBCs is changed only transiently, suitable for short-term treatments. Therefore, further experiments were set up with 6h treatments.

4.1.2 – Iron treatments are not affecting parasites' viability

Next, we asked if the treatments were affecting the parasites viability, i.e, if the parasites were being killed or delayed in their development. This is important given that we aimed to identify iron-dependent changes in gene expression, and to avoid those that are merely due to the delayed or halted parasite development. To address this question, we set up the treatments as previously and after 6 hours the medium was replaced with standard culture medium without any treatment and the parasites were allowed to grow for additional 24 hours. Smears were made at the time of replacing medium and after 24 hours in fresh medium (Fig. 4.2).

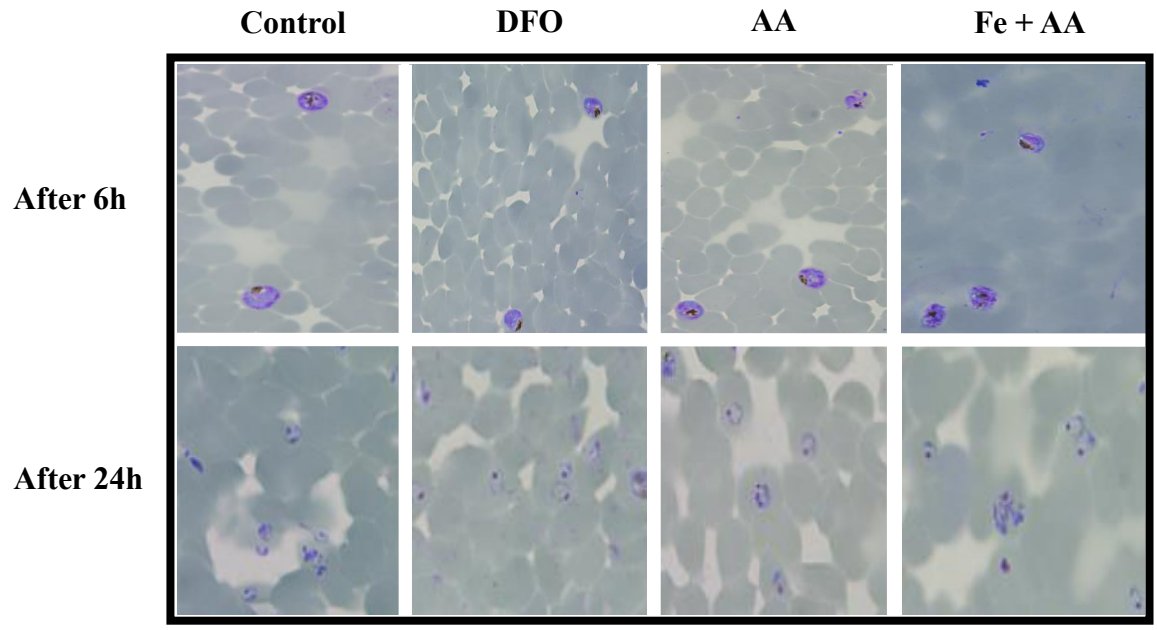


Figure 4.2 – Blood smears of infected RBCs in the different conditions; top row, smears after 6 hours in the treatments, bottom row, smears after 24 hours in replaced medium without any treatment.

As illustrated in figure 4.2, after 6 hours most of the parasites were in the trophozoite stage and we could not observe any difference in the appearance of parasites under treatment compared to control parasites. After 24 hours in fresh medium without the treatments, it seems that the parasites were able to complete the life cycle, fully developing to new merozoites capable of invading new RBCs, as all the conditions had a similar number and developmental stage of ring-stage parasites (parasitemia of 11.4%, 11.5%, 9.3% and 11.4% for Control, DFO, AA and Fe + AA respectively).

This result further supports our system as a valid one to study how the parasites respond to iron without compromising their survival. The next obvious question is if the parasites are in fact responding to these iron perturbations by changing their transcriptome and if we can identify iron-responsive genes.

4.1.3 – Iron treatments are affecting the expression levels of some candidate genes

After confirming that our system was affecting iron levels without compromising parasite's development or survival, we set up treatments in sorbitol-synchronized parasites in the old-ring/young trophozoite stage, and extracted the RNA after 6 hours in order to quantify gene expression of our candidates identified in Table 4.1 by qPCR (see Methods). We chose this "window of time" of the life cycle as it is the most metabolically active one and we considered that it might exhibit the higher iron responsiveness. We also analysed gene expression in other time-points, namely after 2, 24 and 48 hours

of treatment but we encountered some issues regarding them: as the results from the PhenGreen assay showed, the $FeSO_4$ treatment is not inducing changes in the labile iron pool after 24 hours, which we also assume to be true for the 48-hour treatment. On the other hand, the DFO effect seemed to be stronger after 24 hours when compared to 6 hours, which remains to be tested for 48-hour treatment. Regarding the 2 hour time-point, there are studies testing transcriptional responses to compounds showing that after 2 hours one might only detect a "general stress response" instead of a specific response to the treatment (Gupta et al. 2016). For these reasons, only results from 6h treatments will be discussed here.

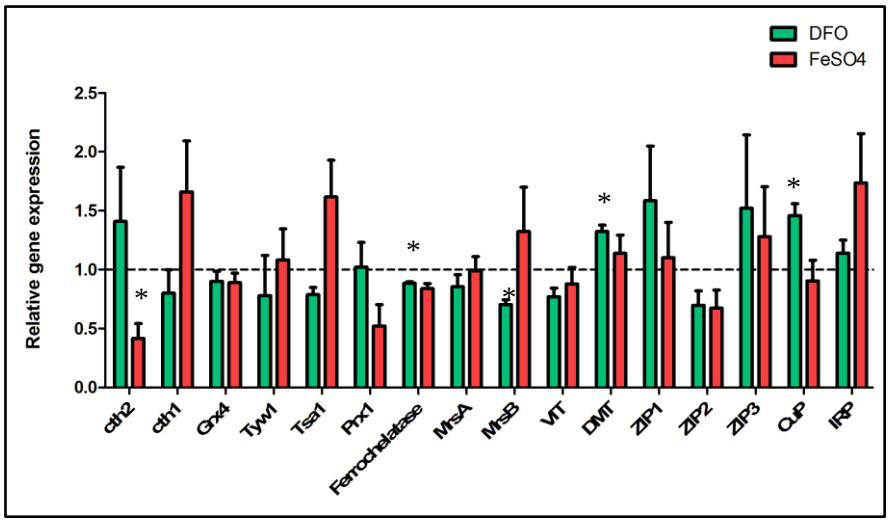


Figure 4.3 – **Relative gene expression in *P. falciparum* after 6 hours with $FeSO_4$ and DFO compared to their corresponding control parasites** Shown results are mean + SEM of 3 independent experiments. One sample t test: * $p < 0{,}005$

Relative gene expression of our candidates determined by qPCR (see Methods for calculation and normalization) after 6 hours in the different iron conditions is illustrated in Figure 4.3. We observed mild changes in expression of this selection of genes, with the highest fold-change around 1.5 and the lowest around 0.5 (dashed line was placed in value 1, i.e, no changes in gene expression). Also, we have observed great variability in expression of some genes between experiments. The lack of reproducibility may be due to the biological variability between experiments, with some genes showing greater changes in one experiment and subtler changes in another, while others were even behaving in opposite directions between experiments. This fact could be explained by the possibility that our candidates do not have the same role as their homologues in other species, meaning they are not responding to changes in iron levels. If this scenario is true, changes in gene expression of some of the candidates are due to stochasticity.

However, several genes showed alterations in expression which were, even though small, reproducible between experiments and statistically significant: starting with DMT, whose expression increases in response to DFO (one sample t-test, p-value = 0.0135). This can point to the role of DMT in re-stablishing normal iron levels inside cells, by transporting it into the parasite cytosol. Ferrochelatase, the last enzyme of the haem biosynthesis pathway, was also significantly down-regulated in DFO condition (one sample t-test, p-value = 0.0184). This result indicates that in iron depletion condition the parasite might be shutting down haem synthesis, one pathway highly dependent of iron. The mitochondrial iron transporter MrsB was also significantly down-regulated in DFO condition (one sample t-test, p-value = 0.0184), indicating reduced iron transport into the mitochondria, as there is less iron available. It is interesting to observe that CuTP, the copper transporter, was significantly up-regulated in the DFO condition (one sample t-test, p-value = 0.0436). This result suggests a potential capacity of this transporter to bind and transport iron, in order to re-establish the normal iron levels inside the cell, or the possibility that some of the copper may be also chelated by DFO. The putative mRNA-binding protein Cth2 was the only one whose gene expression changed in response to Fe significantly (one sample t-test, p-value = 0.0435). Having this result in mind, this gene persists as an interesting target to pursue and unveil its expected mechanism of post-transcription regulation in response to iron levels in the parasite environment.

4.2 – RNA-seq: genome-wide quest for iron-responsive mechanisms

Having the previous results in mind, we decided to advance to the RNA-seq experiment with the objective of genome-wide identification of genes differentially expressed between the iron conditions tested, which we hypothesized to be involved in iron sensing or iron-regulatory processes. The samples for RNA-seq were prepared as described in the Methods section: culture of sorbitol-synchronized old rings/young trophozoites was split into four parallel cultures under the following conditions: 1- Control (regular growth medium); 2- medium with the addition of 200 µM of $FeSO_4$ and 200 µM ascorbic acid; 3 – medium with the addition of 200 µM ascorbic acid and 4 – medium with the addition of 20 µM of the iron chelator deferroxamine (DFO). 6 hours after the start of the treatments, total RNA was extracted to be sent for sequencing, so that at the time of extraction our parasites were all at the trophozoite (24-30 hours post-invasion) (Fig. 4.4). In order to have valid biological replicates, parasites in normal culture conditions were let to complete one cycle (approximately 48 hours) and treatments were repeated. One round of sorbitol was performed in order to keep parasites synchronized. After the RNA extraction was completed for all eight samples, the quantity and quality of RNA was checked. We were able to obtain high amounts of RNA (approximately 20 µg with concentrations above 230 ng/µl for all samples) with good integrity as confirmed by agarose gel and purity as confirmed by Nanodrop measurements (Figure 4.5 and Table 4.2, respectively). The quality of the samples was further confirmed by the company (see Methods).

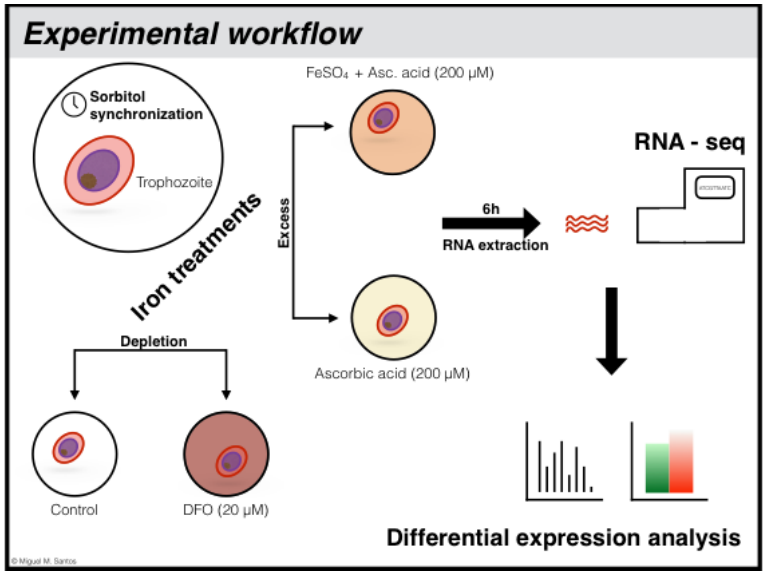


Figure 4.4 – Experimental workflow: sorbitol-synchronised *P. falciparum* blood stage parasites were placed in each condition, RNA was extracted after 6 hours and sent for sequencing in order to identify differentially expressed genes.

Table 4.2 – RNA concentration and ratios measured using Nanodrop and amount of RNA for each sample.

Sample	RNA concentration (ng/µl)	260/280	260/230	Total RNA (µg)
Sample 1 - C1	244.59	2.31	2.50	22.013
Sample 2 - Fe1	308.46	2.17	2.28	27.761
Sample 3 - Aa1	439.01	2.25	2.46	39.51
Sample 4 - DFO1	231.16	2.32	2.72	20.804
Sample 5 - C2	233.9	2.45	2.68	21.051
Sample 6 - Fe2	271.3	2.11	2.03	24.417
Sample 7 - Aa2	242.8	2.45	2.65	21.852
Sample 8 - DFO2	298.1	2.44	2.70	26.829

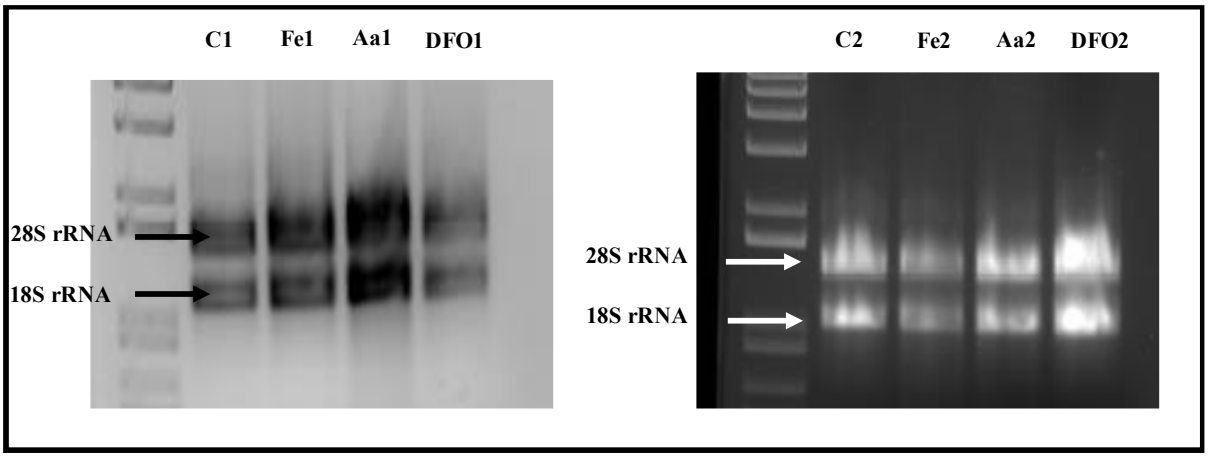


Figure 4.5 - Agarose gels with RNA samples sent for sequencing: left, replicate 1; right, replicate 2, C – Control, Fe – $FeSO_4$ + Ascorbic acid; Aa – Ascorbic acid; DFO – deferoxamine. The bands corresponding to 28S and 18S rRNA are highlighted to confirm RNA integrity.

RNA was sequenced at the STABVIDA facilities through Next-generation Sequencing of the cDNA molecules generated from each sample and raw sequence data of the transcriptome from the total RNA were obtained as FASTQ files. Most samples generated from 25 274 096 to 48 249 776, which is in accordance with the expected output (around 20 000 000 sequence reads). However, the sample 1 (Control 1) had an output below this expectation (14 643 906), which might affect the bioinformatics analyses (see Supplementary information, Table S1). The quality of the data produced was determined using the FastQC tool (see Methods). All the samples had a Phred quality score at each cycle (position in read) indicative of good quality (above 28). The Phred quality score (Q) is a measure of the quality of identification of the nucleobases generated by automated DNA sequencing, which is inversely logarithmically related to the base calling error probabilities (P): $Q = -10 \log_{10}P$. For example, if a base has a Phred score of 10, the probability that it was incorrectly called is 10%. If the score is 20 the probability is 1%, 30 is 0.1%, and so on (Ewing et al. 1998). An example of the plot generated by FastQC tool for Sample 4 - DFO1 and Sample 7 – AA2 can be found in supplementary information (Figure S1).

After mapping the reads to the reference *P. falciparum* 3D7 genome using TopHat (see Methods), we observed that for most samples between 92 and 96% of the reads mapped to the reference genome, except for sample 1 (Control 1) where 87.3% of the reads mapped to the reference genome. This might be due to the lower number of sequence reads originated from this sample.

4.2.1 – A replication issue: when the highly regulated *P. falciparum* blood stage transcriptome becomes a problem

After having the reads from each sample mapped to the reference genome, the transcripts assembled and a merged single reference transcriptome generated from all samples using Cufflinks and Cuffmerge, respectively (see Methods), we proceeded to the differential expression analysis. Since two independent biological replicates were sequenced for each condition, replicates were initially pooled for the differential expression analysis, as it is normally performed in similar studies (Broadbent et al. 2015).

However, after performing the differential expression analysis comparing pooled replicates of Control *vs.* DFO and $FeSO_4$ *vs.* Ascorbic acid conditions, none of the genes was significantly up or down regulated within each pair of conditions. More specifically, for genes which were at least 2-fold up or down-regulated, obtained false discovery rate values (FDR) were not supporting statistical significance.

This result was unexpected given the observed levels of fold-change for the most highly differentially expressed genes (log_2fold change 4.461 and -2.833 for $FeSO_4$ *vs.* Ascorbic acid and 7.20 and -7.51 for Control *vs.* DFO) and thus we hypothesized that the variability between the biological replicates could be the reason for the outcome of the initial analysis. The highly periodic pattern of the blood stage transcriptome means that small shift in the developmental stage of the parasites leads to great difference in gene expression values, as discussed in the Introduction (Lemieux et al. 2009). Therefore, one possibility was that between the two days of the experiments the parasites were not exactly in the same developmental stage, which means that they varied significantly in terms of gene expression and that considered biological replicates were not true replicates. This would mean that the gene expression changes due to the different development stages between the replicates could be "masking" the changes due to the iron treatments.

To test this hypothesis, we performed the differential expression analysis considering the day of the experiment as the variable introducing changes in gene expression levels, meaning the four conditions set up on the first day were replicates of each other, and the same for the conditions set up on the second day. We are sure that the parasites within each day are exactly in the same developmental stage, as each treatment was set up separately from a single culture of sorbitol-synchronized parasites. Also, by pooling the four experiments of each day as replicates we are increasing the statistical power of the analysis, so it would be possible to identify significant differentially expressed genes.

The results obtained by this analysis confirmed our hypothesis, as out of the 5853 genes 300 were significantly up-regulated (log_2 fold change > 1 and p-value = 5,00E-05) and 130 were significantly down-regulated (log_2 fold change < -1.9 and p-value = 5,00E-05) in the samples from second day compared to the first. The first evidence indicating that our parasites were in a later developmental stage on the second day of experiment comparing to the first was the fact that seven SERA (Serine Repeat Antigen) genes were up-regulated on day 2, and it is already described that these subset of genes is up-regulated later in the blood stage (Miller et al. 2002). To address this, we compared our data with the RNA-seq dataset across the blood stage from Otto et. al (Otto et al. 2010) available on PlasmoDB (Fig. 4.6). In their dataset, they identified 1140 genes up-regulated (fold difference > 2) in the 32-hour time-point (corresponding to old trophozoite) compared to the 24 hour time-point (corresponding to young trophozoite). In our dataset, of the 300 identified as being upregulated on the second day compared to the first day, 187 (62%) overlapped with data from Otto et. al, which further strengthens the hypothesis that our parasites were in a later developmental stage in the second day compared to the first (Figure 4.6 left). In addition, in their dataset they identified 696 genes down-regulated (fold difference > 2) in the 32

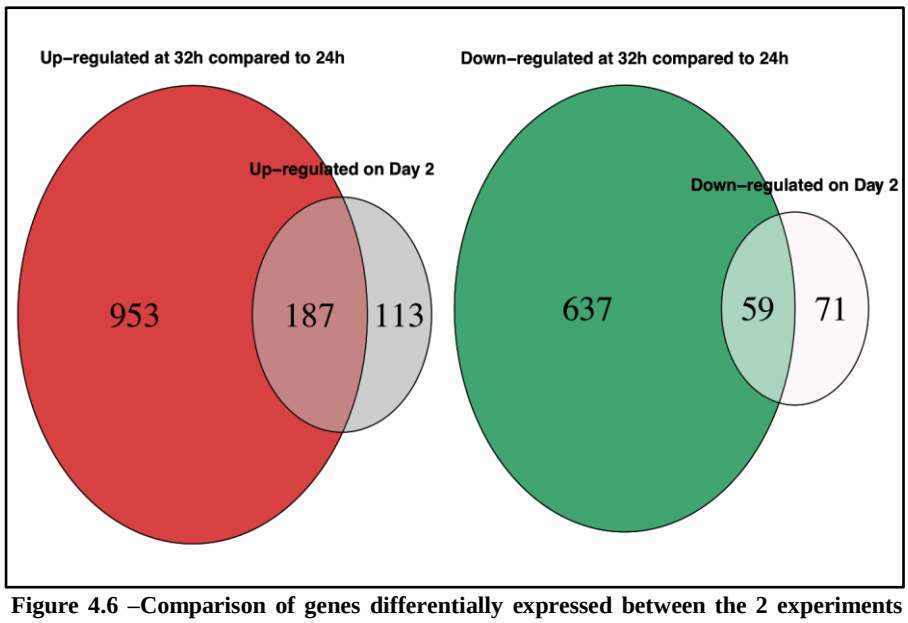


Figure 4.6 –Comparison of genes differentially expressed between the 2 experiments with the data published by Otto et. al confirms the shift in parasite development between experimental days: left: of the 300 up-regulated on the second day, 187 overlapped with the data published; **right**: of the 130 down-regulated on the second day, 59 overlapped with the data published.

hour time-point compared to the 24 hour time-point and in our dataset, of the 130 genes identified as being downregulated on the second day compared to the first 59 (45%) overlapped with data from Otto et. al, validating our hypothesis (Figure 4.6 right).

4.2.2 – The clock-like pattern of gene expression

As we consider that the replication issue we faced is of major importance to further explain all our results, we decided to perform a Multidimensional scaling (MDS) to assess stage similarities on a transcriptome-wide scale (see Methods). This approach was used by Broadbent et al., in their work where they performed RNA-seq across the blood stage in order to identify developmentally regulated lnc-RNA (long non-coding RNAs) (Broadbent et al. 2015). By performing this analysis, they were able to show that their samples were placed around a circle, progressing through the 48 hours *P. falciparum* blood stage according to their time and morphology. We downloaded their gene expression results and performed the same analysis, and were able to obtain the same clock-like pattern (Figure 4.7 - A). Then, the analysis was repeated together with our data, in order to identify where our samples would fall in the "blood-stage transcriptomic clock" (Figure 4.7 - B).

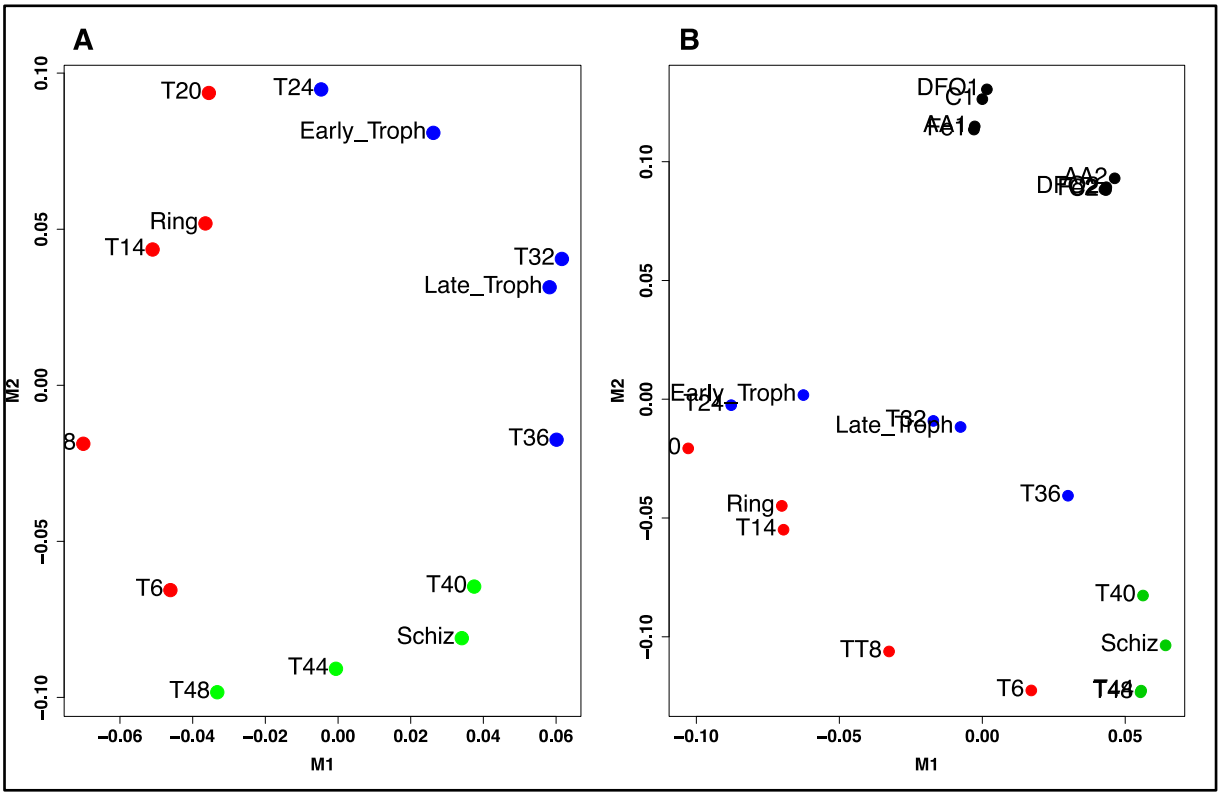


Figure 4.7 – Multidimensional scaling (MDS) confirms the *P. falciparum* transcriptome pattern across the blood stage: **A -** analysis performed with the data from Broadbent et. al; **B** – same analysis with addition of our data. The colour scheme is the following: red corresponds to ring stage, blue to trophozoite and green to schizont, while our data points are coloured in black.

By looking to Figure 4.7-B it is possible to see that the clock pattern is preserved, and that our samples from each day group together with each other. However, the points representing our samples do not fall exactly in the circle, which is expected as our samples were not treated exactly in the same way as the ones from the paper (RNA extraction and the library preparation protocol were not the same, for example). Despite that, a clear shift is present, with samples from day 1 appearing more to the left (younger in development) compared to samples from day 2. This result confirms the assumption that

parasites were in different development stages between the two days of the experiments, suggesting that they should not be analysed as biological replicates.

4.2.3 – Pairwise comparison: finding common patterns between treatments and experiments

After our analysis showed that the data from the two days of experiments should not be considered as replicates, we decided to perform a pairwise comparison between each treatment and its corresponding control (DFO vs. Control and Fe vs. Ascorbic Acid) separately for each experimental day. We aimed to identify genes which responded in an opposite direction between the iron excess and iron depletion conditions in the same day of the experiment. Also, genes behaving in the same way between both days of experiments are more likely to be truly involved in iron-related functions irrespective of the developmental stage of the parasites.

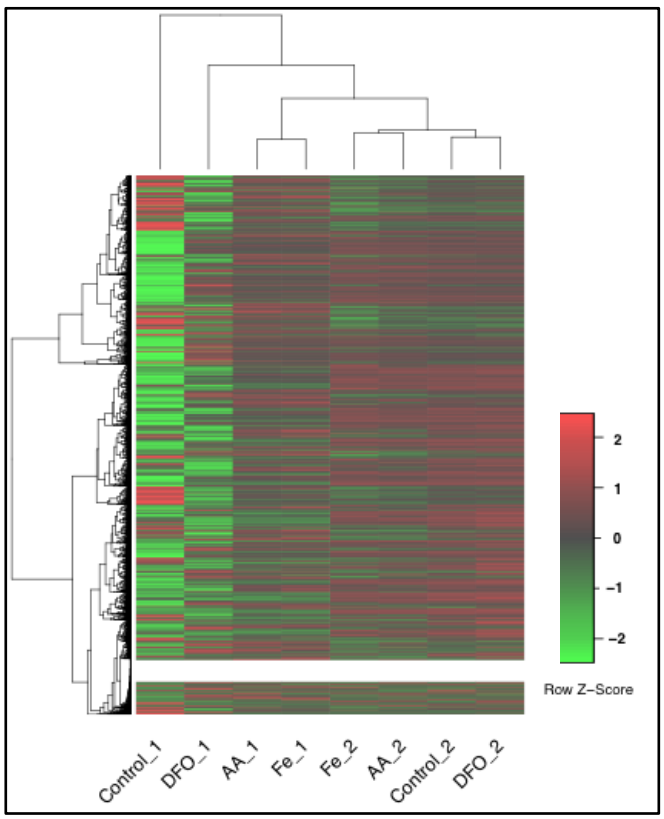


Figure 4.8 - Clustered heatmap shows the visual representation of the samples' comparison based on the expression values of each transcript as log_2(FPKM+1): the four samples from day 2 cluster together and completely apart of samples from day 1. Cluster of genes is based on the mean of log_2(FPKM+1) for each gene.

Figure 4.8 represents a clustered heatmap based on the expression levels of each transcript (see Methods). Samples are clustered based on the mean of log_2(FPKM) of all genes. It is evident that sample 1 (C1) is clearly unrelated from all the other, which might also be a factor affecting the analysis and further supporting pairwise comparisons instead of analysis with pooled "replicates". This difference in C1 sample is likely an artefact due to the lower number of reads generated from this sample. To avoid artefacts in expression changes which are probably result of technical problems during the library preparation or sequencing reaction, we decided to not consider sample 1 (C1) as a valid one, and therefore comparison between Sample 1 (C1) and Sample 4 (DFO1) will not be presented here. The cluster of the genes followed the same rule as the cluster of the samples, i.e., the mean of log_2(FPKM+1).

Since pairwise comparisons are made without replicates, the p-values obtained by the analysis cannot be considered. In order to identify genes with changes in gene expression between the two iron conditions, we made a cut-off at the log_2(fold-change) as follow: a gene is considered up-regulated if the log_2(fold change) is higher than 1 (fold change> 2) and is considered down-regulated if the log_2(fold change) is lower than -1 (fold change<0.5).

Additionally, the FPKM value from each gene was taken into account. Genes for which the FPKM was low to a point that the software was not able to compare the gene expression between the samples (usually <0.1) were not considered for the analysis. It is important to note that in general our results resemble the studies addressing transcriptomic changes in iron deprivation conditions in yeast (see

Introduction), with around 3% of the genes having changes in their expression between different iron conditions.

From the experimental day 1 samples, as explained previously, we only analysed the results from the comparison between ascorbic acid (AA) and $FeSO_4$ (Fe) conditions. Of the 5434 genes, after removing genes which did not have a FPKM value > 0 in both samples, 109 genes were up-regulated and 81 were down-regulated according to our cut-off settings. The gene ontology enrichment was performed but was found not informative as of all differentially expressed genes, 70 were annotated with "unknown function". The remaining genes were carefully investigated in order to identify those which might be iron-related: an ATP synthase mitochondrial F1 complex assembly factor 1 (ATP11, PF3D7_1209800) was up-regulated with Fe treatment (log_2 fold change = 1.35). This result is potentially interesting, as it might indicate that the parasite is reprogramming its metabolism in response to the increase in the iron concentration. One gene encoding a protein with a zinc finger domain (ZF1, PF3D7_1407700) was also in the group of genes up-regulated with the Fe treatment, with a log_2(fold change) of 1.16. These motifs are usually present in transcription factors, and as such this gene may have a role in transcriptional regulation of genes involved in iron metabolism (as it happens in yeast). It is also interesting that ubiquitin-activating enzyme E1 (UbE1, PF3D7_1350400) was up-regulated with the Fe treatments (log_2 fold change = 1.57). This may indicate the existence of iron-responsive proteasomal degradation of proteins involved in iron metabolism is iron-responsive, as observed in the mammalian cells (Thompson & Bruick 2012), which could be further tested through proteomic approaches. Interestingly, a putative magnesium transporter (MgT1, PF3D7_0827700) was also identified as up-regulated with Fe (log_2fold = 2.5), which might suggest promiscuity of this transporter for its substrate.

Analyses of the samples collected on the second experimental day, included both DFO and Fe treatment conditions. For DFO *vs.* Control pairwise comparison, out of the 5424 genes with a FPKM value above > 0 in both samples, 119 were up-regulated and 61 were down-regulated with the DFO treatment. Again, the gene ontology enrichment lacked meaningful and informative GO terms due to the high number of genes annotated as "unknown function" (64 out of the all differentially expressed). We investigated carefully the remaining differentially expressed genes in order to find potential iron-related functions. PUF2 (PF3D7_0417100), an mRNA-binding protein was up-regulated with the DFO treatment (log_2fold change = 1.09). This might indicate a mechanism of post-transcriptional regulation in response to iron deprivation conditions, as the one present in yeast and mammalian cells using Cth1/Cth2 and IRP, respectively. Another interesting result was upregulation (log_2fold change = 1.06) of the RNA of unknown function, RUF6-2 (PF3D7_0711800). Although the FPKM values were low in both samples, we decided to keep it as a candidate to pursue, as non-coding RNAs which are lowly expressed might have important biological functions. A SET domain protein (SET4, PF3D7_0910000) was also found up-regulated with the DFO treatment (log_2fold change = 2.1). This protein has histone-lysine N-methyltransferase activity, and it might be involved in epigenetic mechanisms of gene regulation, by introducing methyl groups in the histones of the chromatin. This would be the first description of the involvement of epigenetic marks in the regulation of iron-related genes in *Plasmodium*. Importantly, the ubiquitin-activating enzyme E1 (UbE1, PF3D7_1350400) described above to be up-regulated with Fe treatment, was found here down-regulated (log_2fold change = -1.02).

Pairwise comparison was also performed between ascorbic acid (Aa) and $FeSO_4$ (Fe) samples collected on the second experimental day: of the 5401 genes with FPKM > 0 in both samples, 89 were up and 111 were down-regulated with the Fe treatment. In this comparison we were able to find several genes potentially involved in iron-related processes: starting with the ABC transporter F family member 1 (ABCF1, PF3D7_0813700) which was up-regulated (log_2fold change = 2.2) in the Fe condition. This

family of transporters have ATPase activity and are involved in drug resistance in *Plasmodium*, as they transport a wide range of structurally and functionally diverse amphipathic compounds (Koenderink et al. 2010). The fact that this gene is up-regulated in an iron excess conditions indicates a possible involvement in iron regulation. Also found up-regulated in the Fe condition (log_2fold change = 1.52) was PF3D7_1467000, a gene assigned as a membrane protein with unknown function, but with assigned GO-terms "magnesium ion transmembrane transporter activity" and "RNA-binding" (referred as MgT2). A folate transporter (FT1, PF3D7_0828600) was down-regulated (log_2fold change = -2.47) in the Fe condition. As there are some evidences that haem transporters and folate transporters might be related (Andrews et al. 2007), we found this observation interesting, as the parasite might be blocking the haem transport whit the increase of iron concentration. Another gene annotated as unknown function (DBP, PF3D7_1231300) was up-regulated (log_2fold change = 1.08) and had the following GO terms: "regulation of transcription, DNA-dependent", "DNA-binding" and "zinc ion binding', which clearly places it has a potential transcription factor. Another transcription factor, belonging to the family of transcription factors with AP2 domains (ApiAP2_1, PF3D7_0404100) was down-regulated in the Fe condition (log_2fold change = -1.17). AP2 family of transcription factors is considered as a major transcriptional regulation in *Plasmodium*, and the AP2 domains they possess are homologous to the AP2 family of transcription regulators in plants, playing important roles in development and environmental stress response pathways (De Silva et al. 2008). It is tempting therefore to propose this specific AP2 proteins as a transcriptional regulator of genes involved in iron metabolism. Another protein annotated as "unknown function" which was up-regulated (log_2fold change = 1.32) had interesting GO terms associated, such as "tetrapyrrole biosynthetic process" and "uroporphyrogen-III synthase activity" (TePy, PF3D7_1247600). These GO terms are strong indicators that this protein is involved in haem biosynthesis, which is also strongly related to the iron metabolism.

Table 4.3 – Summary of the number of genes differentially expressed in the experimental conditions (up and down-regulated) as well as the number of genes annotated as "unknown function" within the differentially expressed genes

		Day 1	Day 2	Annotated as "unknown function"	
				Day 1	Day2
Fe	Up	109	89	70	62
	Down	81	111		
DFO	Up	-	119	-	64
	Down	-	61		

The group of differentially expressed genes upon Fe treatment on day 2, includes genes that are expressed in opposite directions compared to the DFO treatment. This includes RUF6-2, the RNA of unknown function which was up-regulated with DFO. This gene was down-regulated (log_2fold change = -1.31) with the Fe condition. The SET4 protein up-regulated with the DFO treatment was down-regulated (log_2fold change = -1.11) with the Fe, raising the probability of being involved in an iron-related mechanism. Table 4.3 summarizes the number of genes differentially expressed between each conditions on both experiments, as well as the number of genes annotated as "unknown function" (Control *vs.* DFO from day 1 are excluded for the reasons already discussed).

As explained previously, due to the problem of developmental stage difference between the 2 experiments, we considered the pairwise gene expression comparison as the most valid approach. Nevertheless, pooled analysis of 'replicate' experiments may identify genes which are differentially expressed irrespectively of the developmental stage of the parasites. Thus, we performed this analysis for the Ascorbic acid and $FeSO_4$ treated samples (DFO-treated samples could not be analysed in this way given the technical problems with control 1 sample). Using this analysis, there were some

interesting results: expression of two ApiAP2 transcription factors was changed in opposite directions, as one (ApiAP2_2, PF3D7_1317200) was up-regulated and the other (ApiAP2_3, PF3D7_1429200) was down-regulated (log_2fold change = 1.48 and -2.35, respectively). Surprisingly, the latter had the sequence to which the former specifically binds, pointing to activation of a transcriptional cascade in response to changes in the iron levels. Finally, PF3D7_0519300 a cytochrome c oxidase assembly protein (haem A: farnesyltransferase) was up-regulated (log_2fold = 1.5), again showing the possibility that the parasite is regulating its haem biosynthesis pathway in response to iron perturbations.

Table 4.4 – Genes differentially expressed in the different iron conditions: genes are arranged in categories and the fold-changes determined by qPCR and RNA-seq are displayed. The colour code gradient is the following: red up-regulated, green down-regulated, white no change in gene expression. Note that the fold changes determined by RNA-seq are much higher and lower for up and down-regulated genes, respectively. Genes that were determined as up or down-regulated with only one of the treatments only have that value of fold-change and the respective qPCR.

Category	Name	qPCR fold change (Fe/DFO)		RNA-seq fold change (Fe/DFO)	
Transcriptional regulation	ZF1	1,57	-	2,23	-
	ApiAP2_1	0,78	-	0,44	-
	ApiAP2_2	0,85	-	2,79	-
	ApiAP2_3	0,98	-	0,20	-
	DBP	1,02	-	2,11	-
Non-transcriptional regulation	PUF2	-	1,17	-	2,13
	RUF6-2	2,9	0,79	0,40	2,08
	SET4	1,03	1,40	0,46	4,28
	UbE1	1,51	1,19	2,96	0,49
Metabolism	ATP11	1,1	-	2,55	-
	ABCF1	0,85	-	4,59	-
Transport	MgT1	1,33	-	5,66	-
	MgT2	1,02	-	2,87	-
Haem metabolism	FT1	0,75	-	0,18	-
	TePy	0,7	-	2,50	-
	Farnesyltransferase	0,95	-	2,83	-

Finally, to validate our RNA-seq results we designed primers to test a group of differentially expressed genes selected based on the level of the fold change and putative function, by qPCR analysis. The directions of gene expression changes detected by qPCR seemed to follow the trend of RNA-seq results, meaning almost all genes tested were up or down-regulated both in RNA-seq and qPCR results. The genes whose validation by qPCR revealed opposite results were among the ones with the lowest FPKM values, meaning they had less sequencing reads which might compromise the mapping and consequent analysis. Also, the fold change values were found to be much higher or lower for up and down-regulated genes, respectively, in the RNA-seq results compared to the qPCR. This is often reported to occur because the normalization methods are different (FPKM in RNA-seq *vs.* relative expression comparing

to the house-keeping genes in qPCR) and also due to the higher sensitivity of RNA-seq to detect subtler changes. Also, it might be the case that the primers we designed do not allow 100% efficiency of PCR amplification of our target genes, which could introduce changes in the fold change values. Finally, prior to RNA-seq library preparation, samples were depleted for ribosomal RNA, while our qPCR was performed on total parasite RNA, which may also lead to differences in analysis. The information about the genes' category of function and fold changes determined by qPCR and RNA-seq are displayed on table 4.4.

5 – Discussion

As iron is essential for both the host and the pathogen, complex interactions develop between the two organisms that are in a battle for this micronutrient. Consequently, understanding this host-pathogen interaction may become a herculean task. Both the replication of the parasites and the immunity of the host are affected by iron levels (see Introduction). Therefore, many studies investigated the influence of host iron status on the outcome of malaria disease (Clark et al. 2014). However, it is important to note that the most basic questions of how iron directly affects the malaria parasite development remains unknown and under-investigated. Also, fundamental questions of how the *Plasmodium* acquires iron at the different stages during its life cycle and what is the source of iron that it uses, remain open. Most importantly, it is not known how is the signal of the iron levels in host transmitted into the parasite's replication "decision". Our focus was therefore directed to investigate the specific effect of iron perturbations on *P. falciparum* transcriptome.

Throughout its life cycle, *Plasmodium* development is controlled by an amazing gene expression program allowing this unicellular organism to swap between two completely different hosts, travel and adapt to different cell types and to switch from a highly replicating asexual stage to a transmission sexual stage. It is scary and beautiful at the same time how extremely well adapted the parasite is to us; how it turns on and off specific genes accordingly to the most important cellular functions at a particular time of its development. Nothing is done by chance in *Plasmodium.*

Transcriptional factors from the ApiAP2 family are emerging as the critical regulators of gene expression in *Plasmodium*, especially for their role in regulation of transitions between developmental stages, such as transition to form gametocytes, oocysts and liver stages. We identified several transcription factors from ApiAP2 family to be affected by our treatment conditions. An interesting result was that ApiAP2_2 (PF3D7_1317200) was up-regulated in Fe-treated samples, and that we found 3 differentially expressed genes that contained specific binding sites for this transcription factor: another ApiAP2_3 (PF3D7_1429200) which was down-regulated in Fe-treated samples, the mRNA-binding protein PUF2 which was up-regulated with DFO and the histone methyltransferase SET4 which was down and up-regulated in Fe and DFO-treated samples, respectively. These evidences start to place this transcription factor as a potential master regulator, as it is regulating the expression of several genes affected by the Fe treatments, which by themselves regulate gene expression by post-transcriptional mechanisms. It is interesting to note that it is already described that the disruption of this transcription factor leads to the blockage in gametocyte production (Josling & Llinás 2015). It would be interesting to address if is there any link between disruption in iron levels and transcription control of gametocyte production.

In addition to this transcriptional control, the gametocyte production is controlled post-transcriptionally by the RNA-binding protein PUF2 which works as a repressor of gametocyte production (Miao et al. 2010). Therefore, it is interesting that we found PUF2 to be up-regulated with the DFO treatment. This provides another evidence supporting a relationship between iron levels and gametocyte production. Recently, the importance of translational regulation in gametocytes was discovered: by combining transcriptomic and proteomic data, Lasonder et al. observed that 512 highly expressed transcripts in the female gametocyte did not have the corresponding protein expression, showing a large scale translational repression in *P. falciparum* female gametocytes (Lasonder et al. 2016). Surprisingly, for some of these transcripts, the protein expression was only detected in oocysts and sporozoites, showing that repressed transcripts can be silenced for both short and long time periods. Based on our PUF2 result,

it would be of interest to further investigate the effect of iron on the translational regulation of sexual development of malaria parasites.

Also, a component of the ubiquitin-proteasome system was affected by both the iron depletion and excess conditions, showing that the parasite might also be using this pathway as a way of tackling this stress situation. Iron-dependent degradation by the ubiquitin-proteasome pathway was already described as major player in the cellular iron homeostasis in human cells, through affecting the levels of IRP1 and IRP2, ferroportin and Hypoxia Inducible Factor (HIF) (Wang et al. 2007; Thompson & Bruick 2012; Vashisht et al. 2009). Therefore, it would be interesting to perform proteomic analysis of *Plasmodium* parasites developing in different iron conditions to gain better understanding of putative post-transcriptional and post-translational iron regulatory mechanisms.

An interesting finding from our RNA-seq analysis is change in expression of RUF6, a RNA of unknown function which was also affected by both conditions but in opposite directions. Is there the possibility that non-coding RNAs are involved in iron sensing and regulation? This was already observed in the cyanobacteria *Synechostis,* where several ncRNAs were identified as differentially expressed after an iron deprivation condition (Hernández-Prieto et al. 2012). In mice, it was shown that one liver-specific microRNA is involved in the regulation of expression of hepcidin mRNA and iron levels in the tissues (Castoldi et al. 2011). Thus, investigation of this RNA molecule in *P. falciparum* deserves further experimentation to understand its putative role in regulation of iron metabolism. Moreover, the expression of the histone-methyl transferase SET4 was affected by both iron conditions, shedding the light on a possible epigenetic mechanism of transcriptional regulation in response to iron, as it happens in *Arabidopsis* (Xing et al. 2015). It would be interesting to pursue which specific genes are target of this methylation.

In addition to above mentioned putative gene expression regulators, several putative transporter genes and genes encoding enzymes involved in iron-consuming metabolic pathways, were also found differentially expressed between different iron treatments. The changes in expression of ATP11, a mitochondrial ATP synthase, highlight the possibility of some kind of metabolism reprograming. Additionally, it was already described that these proteins isolated from rat hearts have iron transport activity (Kim & Song 2010). Finally, haem transport and biosynthesis seem to play an important role in this process, as we identified two proteins involved in haem biosynthesis and a folate transporter (might be transporting haem?). It was already described the importance of haem synthesis, transport and degradation to hemozoin in *Plasmodium* (see Introduction, *Iron in Plasmodium: what do we know?*).

It seems that the parasite not only adapt by changing its metabolism and transport system to overcome the changes in iron levels as we expected, but that it also reprograms its gene and protein expression systems. Thus, *P. falciparum* response to iron perturbations likely involves extremely complex and tightly regulated network of pathways that act downstream on a broad variety of target proteins (Fig 5.1). As in other organisms, iron must be carefully regulated in order to be kept at a concentration high enough for the fulfilling of the cellular processes where it is involved, and avoid the formation of reactive oxygen species and toxicity by iron overload.

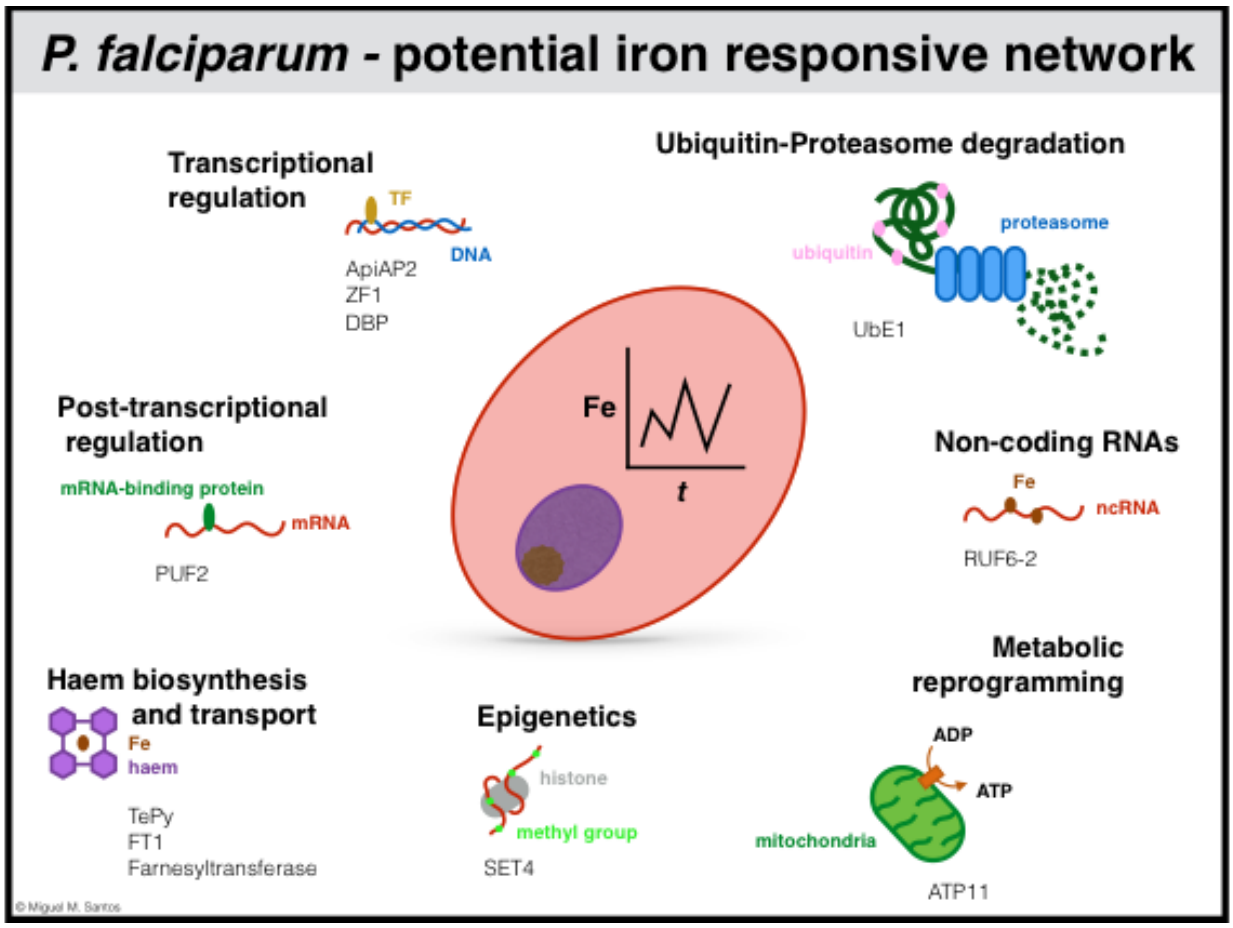


Figure 5.1 – *Plasmodium falciparum* putative iron-responsive network: our findings indicate that the parasites does not rely on a simple alteration in the expression levels of iron transports or proteins directly involved in iron but in a broader system including transcriptional, post-transcriptional and epigenetic regulation, as well as ncRNAs, metabolism, haem biosynthesis and proteasome dependent protein degradation. It is undeniable that the parasite has a highly complex network of iron regulation, involving several layers of control, showing the importance of the maintenance of iron homeostasis for *P. falciparum.*

Regarding our approach, it is important to highlight the fact that by qPCR (and RNA-seq) we are measuring mRNA levels. Since the beginning of this work, we highlighted the importance of the post-transcriptional layer of regulation in gene expression. It might be the case that for many genes their mRNA levels are not significantly altered between the different iron conditions, but that their translational efficacy is affected or that they are regulated at the protein-level. Therefore, integration of RNA-seq studies with other techniques such as proteomics would be useful in order to fill the most gaps possible and increase the probability of reaching valid conclusions.

Some challenges were also faced during this work: it is obvious that when studying *P. falciparum* transcriptome, one must be aware of the strictly developmental regulation of gene expression. If the purpose is to study a specific stage during the intra-erythrocytic development, one must be completely sure that the parasites are exactly on the same stage between the biological replicates (since even 1-2 hour differences can represent a great change in overall gene expression) by performing several rounds of synchronization. The safest approach might be to start parallel cultures from one initial synchronized culture and study the transcriptome on different time-points during development (as the studies from Llinas, Bozdech and Otto described in the Introduction). The major issue with this approach is the cost of RNA-seq analysis. As the sequencing protocols are becoming a common approach and essential part of almost every scientific project, the drop in the cost will in future allow the increase in the number of replicates and conditions. These approaches are important to identify missing genes involved in iron

regulation and metabolism in *P. falciparum* which do not share a substantial percentage of sequence identity with iron-responsive genes in other organisms.

Approximately half of *P. falciparum* genes possess introns, making an isoform differential expression analysis in the different iron conditions an attractive approach. This would show if the alternative splicing is another mechanism involved in iron homeostasis, as it is in proliferation inside RBCs (Eshar et al. 2012). Ideally, we would have proteomic data to complement and increase the reliability of our findings. Also, we consider that if we analyse the transcriptome response to our iron treatments in different time points of the blood stage, we would be able to see how the iron regulation is coordinated with the life cycle progression. Experiments in liver stage would be also useful, as the iron metabolism of the parasite at this stage is unknown and highly under-investigated.

Several interesting studies could be performed based on our findings, that would increase our understanding of the iron biology of the parasite, and consequently in its basic biology: it would be of great relevance to compare the transcriptome of *P. falciparum* parasites collected from anaemic and non-anaemic patients (possibly resembling our iron depletion and excess condition, respectively). Another highly informative experiment would be to put the blood of these patients on culture with our iron treatments and study the transcriptome before and after the treatments, to observe the adaptation of the parasite to the different iron environments and try to revert the gene expression program only with iron fluctuations. If we were able to understand the parasite iron adaptation and predict when and how it will do it, we will be able to design more powerful therapies, combining drug administration with iron supplementation.

The work developed throughout this thesis had the major objective of ironing out the mechanisms behind iron homeostasis during the blood stage of *P. falciparum* parasites. Obviously, it represents a huge challenge which should be approached from several directions and using multiple strategies. However, we were able to observe that blood stage *P. falciparum* responds to environmental iron changes by mounting a transcriptomic response. The analysis of this response indicates that this parasite does not rely on a single mechanism of iron regulation, but that in fact a highly complex network of pathways act on different fronts, starting with the gene and protein expression regulation and extending to metabolic pathways. We consider that we were able to broaden the horizons of the iron sensing and regulation in the malaria parasite, and opened doors for multiple new hypothesis which are relevant to pursue.

References

Aisen, P., Enns, C. & Wessling-Resnick, M., 2001. Chemistry and biology of eukaryotic iron metabolism. *International Journal of Biochemistry and Cell Biology*, 33(10), pp.940–959.

Andrews, N.C. et al., 2007. When Is a Heme Transporter Not a Heme Transporter? When It's a Folate Transporter. *Cell Metabolism*, 5(1), pp.5–6.

Anon, 2015. Who Report 2015. Available at: http://www.who.int/entity/malaria/publications/world-malaria-report-2015/report/en/.

Beinert, H. & Kennedy, M.C., 1993. Aconitase, a two-faced protein: enzyme and iron regulatory factor. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 7(15), pp.1442–1449.

Bozdech, Z. et al., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biology*, 1(1), pp.85–100.

Broadbent, K.M. et al., 2015. Strand-specific RNA sequencing in Plasmodium falciparum malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC genomics*, 16(1), p.454.

Bunnik, E.M. et al., 2016. The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. *Genome Biology*, 17(1), p.147.

Caro, F. et al., 2014. Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *eLife*, 3, pp.1–24.

Castoldi, M. et al., 2011. The liver-specific microRNA miR-122 controls systemic iron homeostasis in mice. *The Journal of clinical investigation*, 121(4), pp.1386–96.

Chu, Y. & Corey, D.R., 2012. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic acid therapeutics*, 22(4), pp.271–4.

Clark, M. a., Goheen, M.M. & Cerami, C., 2014. Influence of host iron status on Plasmodium falciparum infection. *Frontiers in Pharmacology*, 5(May), pp.1–12.

Coulson, R., Hall, N. & Ouzounis, C., 2004. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome research*, pp.1–7.

Craig, A.G. et al., 2012. The role of animal models for research on severe malaria. *PLoS pathogens*, 8(2), p.e1002401.

De Domenico, I., McVey Ward, D. & Kaplan, J., 2008. Regulation of iron acquisition and storage: consequences for iron-linked disorders. *Nature reviews. Molecular cell biology*, 9(1), pp.72–81.

Drakesmith, H. & Prentice, A.M., 2012. Hepcidin and the Iron-Infection Axis. *Science*, 338(6108), pp.768–772.

EM, B. et al., 2014. Polysome profiling reveals translational control of gene expression in the human malaria parasite Plasmodium falciparum. *Genome biology*, 14, pp.1–18.

Eshar, S. et al., 2012. A novel Plasmodium falciparum SR protein is an alternative splicing factor required for the parasites' proliferation in human erythrocytes. *Nucleic acids research*, 40(19), pp.9903–16.

Ewing, B. et al., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research*, 8(3), pp.175–85.

Ferrer, P. et al., 2012. Antimalarial iron chelator, FBS0701, shows asexual and gametocyte Plasmodium falciparum activity and single oral dose cure in a murine malaria model. *PLoS ONE*, 7(5), pp.1–7.

Flo, T.H. et al., 2004. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature*, 432(7019), pp.917–921.

Gardner, M. et al., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419, pp.498–511.

Griffith, M. et al., 2010. Alternative expression analysis by RNA sequencing. *Nature methods*, 7(10), pp.843–7.

Gunshin, H. et al., 1997. Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature*, 388(July), pp.482–488.

Gupta, D.K. et al., 2016. DNA damage regulation and its role in drug-related phenotypes in the malaria parasites. *Scientific reports*, 6, p.23603.

Hentze, M.W. et al., 2010. Two to Tango: Regulation of Mammalian Iron Metabolism. *Cell*, 142(1), pp.24–38.

Hernández-Prieto, M.A. et al., 2012. Iron deprivation in Synechocystis: inference of pathways, non-coding RNAs, and regulatory elements from comprehensive expression profiling. *G3 (Bethesda, Md.)*, 2(12), pp.1475–95.

Hoeijmakers, W.A.M., Bártfai, R. & Stunnenberg, H.G., 2013. Transcriptome Analysis Using RNA-Seq. In *Malaria: Methods and Protocols*. pp. 221–239.

Hu, G. et al., 2010. Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. *Nature biotechnology*, 28(1), pp.91–98.

Illumina, 2015. An Introduction to Next-Generation Sequencing Technology. , (illumina), pp.1–16.

Jo, W.J. et al., 2009. Novel insights into iron metabolism by integrating deletome and transcriptome analysis in an iron deficiency model of the yeast Saccharomyces cerevisiae. *BMC genomics*, 10(1), p.130.

Josling, G.A. & Llinás, M., 2015. Sexual development in Plasmodium parasites: knowing when it's time to commit. *Nature reviews. Microbiology*, 13(9), pp.573–87.

Kenthirapalan, S. et al., 2014. Copper-transporting ATPase is important for malaria parasite fertility. *Molecular Microbiology*, 91(2), pp.315–325.

Kim, M. & Song, E., 2010. Iron transport by proteoliposomes containing mitochondrial F(1)F(0) ATP synthase isolated from rat heart. *Biochimie*, 92(4), pp.333–42.

Koenderink, J.B. et al., 2010. The ABCs of multidrug resistance in malaria. *Trends in parasitology*, 26(9), pp.440–6.

Labbé, S., Khan, M.G.M. & Jacques, J.F., 2013. Iron uptake and regulation in Schizosaccharomyces pombe. *Current Opinion in Microbiology*, 16(6), pp.669–676.

Lasonder, E. et al., 2016. Integrated transcriptomic and proteomic analyses of P. falciparum gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic acids research*, 44(13), pp.6087–101.

Lemieux, J.E. et al., 2009. Statistical estimation of cell-cycle progression and lineage commitment in Plasmodium falciparum reveals a homogeneous pattern of transcription in ex vivo culture. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), pp.7559–7564.

Li, A. et al., 2011. Yap5 protein-regulated transcription of the TYW1 gene protects yeast from high iron

toxicity. *Journal of Biological Chemistry*, 286(44), pp.38488–38497.

Li, L. et al., 2012. A role for iron-sulfur clusters in the regulation of transcription factor Yap5-dependent high iron transcriptional responses in yeast. *Journal of Biological Chemistry*, 287(42), pp.35709–35721.

Li, L. et al., 2010. Genetic dissection of a mitochondria-vacuole signaling pathway in yeast reveals a link between chronic oxidative stress and vacuolar iron transport. *Journal of Biological Chemistry*, 285(14), pp.10232–10242.

Llinas, M. et al., 2006. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Research*, 34(4), pp.1166–1173.

Loyevsky, M. et al., 2003. Expression of a recombinant IRP-like Plasmodium falciparum protein that specifically binds putative plasmodial IREs. *Molecular and Biochemical Parasitology*, 126(2), pp.231–238.

Maher, C.A. et al., 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458(7234), pp.97–101.

Martínez-Pastor, M.T. et al., 2013. Post-transcriptional regulation of iron homeostasis in saccharomyces cerevisiae. *International Journal of Molecular Sciences*, 14(8), pp.15785–15809.

Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1), pp.31–46.

Miao, J. et al., 2010. The Puf-family RNA-binding protein PfPuf2 regulates sexual development and sex differentiation in the malaria parasite Plasmodium falciparum. *Journal of cell science*, 123(Pt 7), pp.1039–1049.

Miller, L.H. et al., 2013. Malaria biology and disease pathogenesis: insights for new treatments. *Nature Medicine*, 19(2), pp.156–167.

Miller, S.K. et al., 2002. A Subset of Plasmodium falciparum SERA Genes Are Expressed and Appear to Play an Important Role in the Erythrocytic Cycle. *Journal of Biological Chemistry*, 277(49), pp.47524–47532.

Moyerbrailean, G.A. et al., 2015. A high-throughput RNA-seq approach to profile transcriptional responses. *Scientific Reports*, 5, p.14976.

Nagaraj, V.A. et al., 2009. Mitochondrial localization of functional ferrochelatase from Plasmodium falciparum. *Molecular and Biochemical Parasitology*, 168(1), pp.109–112.

Neilands, J.B., 1995. Siderophores - Structure and Function of Microbial Iron Transport Compounds. *Journal of Biological Chemistry*, 270(45), pp.26723–26726.

Oshlack, A., Robinson, M.D. & Young, M.D., 2010. From RNA-seq reads to differential expression results. *Genome biology*, 11(12), p.220.

Otto, T.D. et al., 2010. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Molecular Microbiology*, 76(1), pp.12–24.

Outten, C.E. & Albetel, A., 2013. Iron sensing and regulation in Saccharomyces cerevisiae : Ironing out the mechanistic details. *Current Opinion in Microbiology*, 16(6), pp.662–668.

Painter, H.J., Campbell, T.L. & Llinás, M., 2012. The Apicomplexan AP2 family: Integral factors regulating Plasmodium development. *Mol Biochem Parasitol.*, 176(1), pp.1–7.

Portugal, S. et al., 2011. Host-mediated regulation of superinfection in malaria. *Nature Medicine*, 17(6), pp.732–737.

Portugal, S., Drakesmith, H. & Mota, M.M., 2011. Superinfection in malaria: Plasmodium shows its

iron will. *EMBO reports*, 12(12), pp.1233–1242.

Reddy, B.P.N. et al., 2015. A bioinformatic survey of RNA-binding proteins in Plasmodium. *BMC genomics*, 16(1), p.890.

Le Roch, K.G. et al., 2004. Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle. *Science*, 301(5639), pp.1503–1508.

Le Roch, K.G. et al., 2004. Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. *Genome research*, 14(11), pp.2308–18.

Rovira-Graells, N.N. et al., 2012. Transcriptional variation in the malaria parasite Plasmodium falciparum. *Genome Research*, 22, pp.925–938.

Sahu, T. et al., 2014. ZIPCO, a putative metal ion transporter, is crucial for Plasmodium liver-stage development. *EMBO Molecular Medicine*, 6(11), pp.1387–1397.

Sanchez, M. et al., 2011. Iron regulatory protein-1 and -2: Transcriptome-wide definition of binding mRNAs and shaping of the cellular proteome by iron regulatory proteins. *Blood*, 118(22), pp.168–180.

Schade, A.L. & Caroline, L., 1944. Raw hen egg white and the role of iron in growth inhibition of Shigella dysenteriae, Staphylococcus aureus, Escherichia coli and Sac- charomyces cerevisiae. *Science*, 100, pp.14–15.

Shakoury-Elizeh, M. et al., 2004. Transcriptional Remodeling in Response to Iron Deprivation in Saccharomyces cerevisiae □ D. *Molecular Biology of the Cell*, 15, pp.1233–1243.

Shaw, P.J. et al., 2015. Plasmodium parasites mount an arrest response to dihydroartemisinin, as revealed by whole transcriptome shotgun sequencing (RNA-seq) and microarray study. *BMC Genomics*, 16(1), p.830.

Sigala, P.A. et al., 2012. Direct tests of enzymatic heme degradation by the malaria parasite Plasmodium falciparum. *The Journal of biological chemistry*, 287(45), pp.37793–807.

De Silva, E.K. et al., 2008. Specific DNA-binding by Apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences*, 105(24), pp.8393–8398.

Siwo, G.H. et al., 2015. An integrative analysis of small molecule transcriptional responses in the human malaria parasite Plasmodium falciparum. *BMC Genomics*, 16(1), p.1030.

Slavic, K. et al., 2016. A vacuolar iron-transporter homologue acts as a detoxifier in Plasmodium. *Nature Communications*, 7(May 2015), pp.1–10.

Stahel, E. et al., 1988. Iron chelators: In vitro inhibitory effect on the liver stage of rodent and human malaria. *American Journal of Tropical Medicine and Hygiene*, 39(3), pp.236–240.

Thompson, J.W. & Bruick, R.K., 2012. Protein degradation and iron homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(9), pp.1484–1490.

Trager, W. & Jensen, J.B., 1976. Human malaria parasites in continuous culture. *Science (New York, N.Y.)*, 193(4254), pp.673–5.

Trapnell, C. et al., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3), pp.562–78.

Vashisht, A.A. et al., 2009. Control of iron homeostasis by an iron-regulated ubiquitin ligase. *Science (New York, N.Y.)*, 326(5953), pp.718–21.

Vembar, S.S. et al., 2015. The PfAlba1 RNA-binding protein is an important regulator of translational timing in Plasmodium falciparum blood stages. *Genome biology*, 16(1), p.212.

Wang, J. et al., 2015. Haem-activated promiscuous targeting of artemisinin in Plasmodium falciparum. *Nature Communications*, 6, pp.1–11.

Wang, J. et al., 2007. Iron-dependent degradation of apo-IRP1 by the ubiquitin-proteasome pathway. *Molecular and cellular biology*, 27(7), pp.2423–30.

Wang, J. & Pantopoulos, K., 2011. Regulation of cellular iron metabolism. *Biochemical Journal*, 434(3), pp.365–381.

Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63.

Weinberg, E.D., 2009. Iron availability and infection. *Biochimica et Biophysica Acta - General Subjects*, 1790(7), pp.600–605.

Wilkinson, N. & Pantopoulos, K., 2014. The IRP/IRE system in vivo: insights from mouse models. *Frontiers in pharmacology*, 5, p.176.

Xie, S.C. et al., 2016. Haemoglobin degradation underpins the sensitivity of early ring stage Plasmodium falciparum to artemisinins. *Journal of Cell Science*, 129, pp.406–416.

Xing, J., Wang, T. & Ni, Z., 2015. Epigenetic regulation of iron homeostasis in Arabidopsis. *Plant signaling & behavior*, 10(12), p.e1064574.

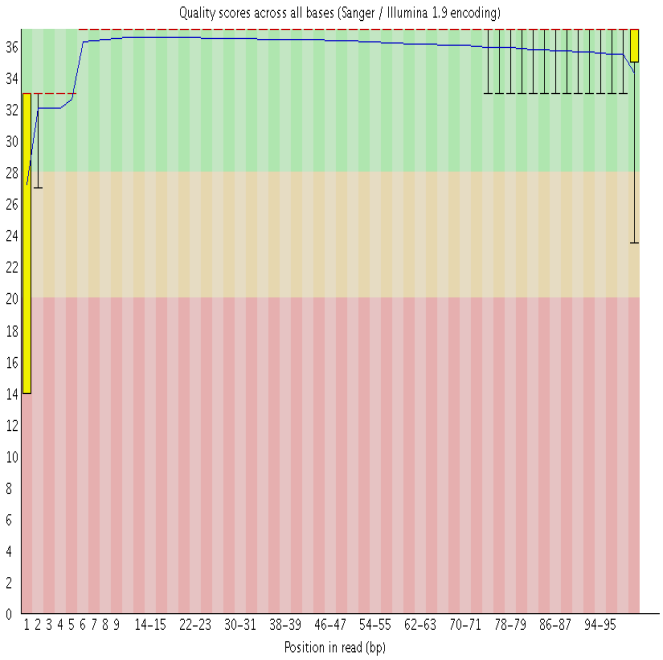
Supplementary information

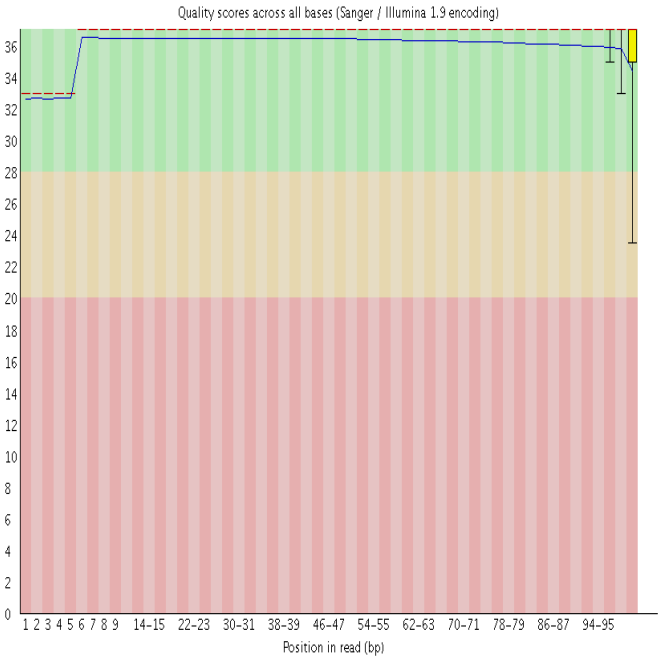
Table S1 – Raw sequence data statistics of each sample.

Sample ID	Number of bases (Mbp)	Number of reads	Mean read length
Sample 1 - C1	1460	14 643 906	101
Sample 2 - Fe1	3290	30 290 286	101
Sample 3 - Aa1	2530	25 274 096	101
Sample 4 - DFO 1	4751	48 249 776	101
Sample 5 - C2	2686	26 872 596	101
Sample 6 - Fe2	2591	25 935 940	101
Sample 7 - Aa2	2586	25 901 590	101
Sample 8 - DFO2	2587	25 895 234	101

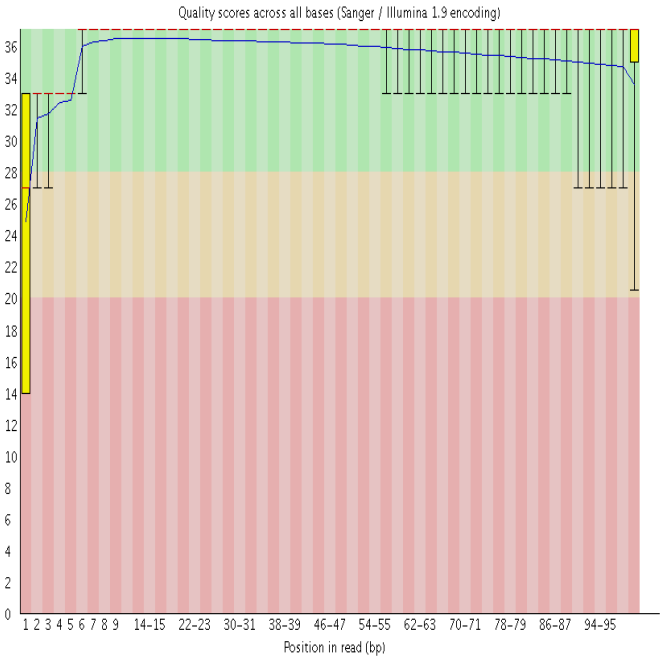
Forward read

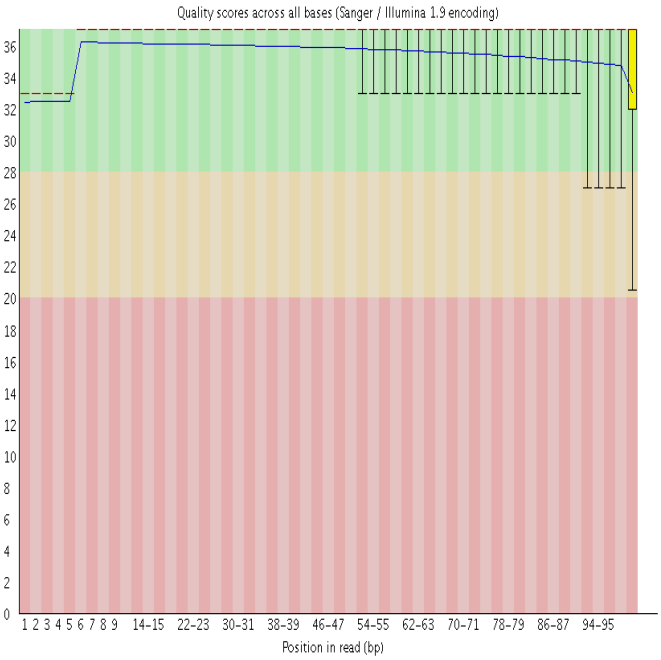
Reverse read





DFO1





AA2

Figure S1 – Plot containing the average quality at each cycle. Top row, sample 4 DFO1; Bottom row, sample 7 AA2; Yellow box: interquartile range (25-70%) of Phred score at each cycle; Red line: median of Phred score at each cycle; Blue line: average of Phred score at each cyle; Green background: good quality (Phred score > 28); Yellow background: acceptable quality (20 < Phred score < 28); Red background: bad quality (Phred score < 20); x-axis: number of cycles; y-axis: Phred quality score

Table S2 – List of primers used for qPCR.

		Gene	P. falciparum ID
Forward	TGTTCATACGCTCATTCACC	Cth2	PF3D7_1134600
Reverse	GTGCCCTTAATTCTAGTTCTCC		
Forward	GAGGTGTTACATAGTAGAACGC	Cth1	PF3D7_0906600
Reverse	TCATTATTACTACTTTCCCTTCCC		
Forward	CTCAATGTAAATTTAGTAATGCC	Grx4	PF3D7_0606900
Reverse	ATATAATTGTGGATATGTAGGC		
Forward	TATCAATGTATGGAAGCTACGC	Tyw1	PF3D7_0524900
Reverse	TGCTTTATCCATATTCCATCTCC		
Forward	CATTAGATAAGGCTCTTGATGC	Tsa1	PF3D7_1438900
Reverse	ATGTATGTTGGATATTTCCTATACC		
Forward	CACTGAACTTGCTGAATTTGG	Prx1	PF3D7_0802200
Reverse	AATATCCCACTTATCTAGGTTTCC		
Forward	ATCACAGGAGCAATTTCTGG	Mrs3	PF3D7_1241600
Reverse	CTCGGATCATAAATGGAGAAGG		
Forward	AAGACAAGAGTCCTCACC	Mrs4	PF3D7_0905200
Reverse	CTATACCGCCACCTATTCC		
Forward	TGGTATGAGATATGGAGAAAGG	Ferrochelatase	PF3D7_1364900
Reverse	CTGCTGATTGTGGATATAATGG		
Forward	TCCGCCTATGTTGCATATAC	VIT	PF3D7_1223700
Reverse	GTAAATTGTGACTTGAACAATCC		
Forward	TGGACATGTGACTGGACTTGA	DMT	PF3D7_0523800
Reverse	CTTGAATATGAGCACCCCAAA		
Forward	GACACTCTTTCACAACATATAACA	ZIP1	PF3D7_0609100
Reverse	GAATAGCCAAAGAGATGGTTA		
Forward	GAAGGCTTACTTATGGGAAG	ZIP2	PF3D7_1022300
Reverse	GCTTGCTAGTAAAGGATTTTCTA		
Forward	GTTAGCAAATATATTTATGGCA	ZIP3	PF3D7_0715900
Reverse	GTTGAAAAATTCTTCTTAACACT		
Forward	CAAATCGGCCCTCTCTTA	CuTP	PF3D7_0904900
Reverse	GTATCTTCCGTTTTTGGCA		
Forward	GCTCGTTATCACCAGGTTC	IRP	PF3D7_1342100
Reverse	CTTCATCTAGGTGCCCACT		
Forward	GGAAACCTTGTTACGACTTCTCC	18S rRNA	PF3D7_1148600
Reverse	TGCTATGCTGCCTGCTCTTA		
Forward	AAGTAGCAGGTCATCGTGGTT	Serine tRNA ligase	PF3D7_0717700
Reverse	TTCGGCACATTCTTCCATAA		
Forward	GACATAAGTTTAGTAGGTCG	Inositol 5-phosphatase	PF3D7_0802500

Reverse	TTCTGACTCCACATCATTTG		
Forward	GGAAATCCATAGACAAACAATG	Adenyllocusccinate lyase (ASL)	PF3D7_0206700
Reverse	TCCTGTGAGAAGTGCTCCAC		
Forward	GTTGCTGTCATCCAATAA	ZF1	PF3D7_1407700
Reverse	TGGGAAGTTTCGGTAAT		
Forward	TTCGTACGGTTTACATAAC	ApiAP2_1	PF3D7_0404100
Reverse	CCATATCCTGTGTATCATATC		
Forward	GTTTCCATAAACCTGATTTG	ApiAP2_2	PF3D7_1317200
Reverse	CTACACCCTTATGATCTTTC		
Forward	AACGGAGAAATGGAAGA	ApiAP2_3	PF3D7_1429200
Reverse	CCTTCCTGTAACTTCTCTA		
Forward	AGAACCATACGTTCCTC	DBP	PF3D7_1231300
Reverse	TTTGGACAAGCACAAAG		
Forward	GTGTAGAAGTATTACCATGTAG	PUF2	PF3D7_0417100
Reverse	GTTGCTCGTTTCCTATTT		
Forward	GTCGTTAGGTATGTAACCTTTC	RUF6-2	PF3D7_0711800
Reverse	GGGTAACTCAGCAGCTC		
Forward	AACGAAACGCACATAAA	SET4	PF3D7_0910000
Reverse	GTGTATGCTATATCAGAAGG		
Forward	ACAAGCCATTCTCTATAAC	UbE1	PF3D7_1350400
Reverse	GCCAGTGATTCAATAATTC		
Forward	CCTTGTTCTCGTGATTTA	ATP11	PF3D7_1209800
Reverse	CTTTCTAATAAGGGCAACT		
Forward	GAGAGTTATTCGAGTGATAC	ABCF1	PF3D7_0813700
Reverse	GTTCTGTTCCGTTTGTT		
Forward	GTATCCTACCCAGACATAA	MgT1	PF3D7_0827700
Reverse	ATACACCATTCTGCTTTG		
Forward	GGAAGGTTCTTACAATCTAT	MgT2	PF3D7_1467000
Reverse	TTGCTCTTTCCTCTTCT		
Forward	CTTCTCTTAGCATCGTTATG	FT1	PF3D7_0828600
Reverse	TCGATATCCGAAGATTGG		
Forward	TCATCATCTATAAGCACATC	TePy	PF3D7_1247600
Reverse	CTTGTTGGGTATCGTAAC		
Forward	CTTTAACTTCCTTTCTAGGT	Farnesyltransferase	PF3D7_0519300
Reverse	GATCCTACAATTGATCCTAC		